GEOMETRY AND ARITHMETIC OF VERBAL DYNAMICAL SYSTEMS ON SIMPLE GROUPS

TATIANA BANDMAN, FRITZ GRUNEWALD, AND BORIS KUNYAVSKIĬ
WITH AN APPENDIX BY NATHAN JONES

Abstract. We study dynamical systems arising from word maps on simple groups. We develop a geometric method based on the classical trace map for investigating periodic points of such systems. These results lead to a new approach to the search of Engel-like sequences of words in two variables which characterize finite solvable groups. They also give rise to some new phenomena and concepts in the arithmetic of dynamical systems.

ἐν ἀρχῇ ἦν ὁ λόγος...

KATA ΙΩΑΝΝΗΝ 1:1[1]

Contents

[1]In the beginning was the Word... John 1:1

1. Introduction

The initial goal of the present paper was to get deeper understanding of what is behind recent results achieved in describing the class of finite solvable groups by identities in two variables [BGGKPP1], [BGGKPP2], [BWW]. Although the results were purely group-theoretic, it was clear that the key role is played by geometry and dynamics. Byproducts of this investigation seem to us not less interesting than the initial problem.

We reformulated the original problem in the language of a verbal dynamical system on an algebraic group G (the notion of its own interest). We study these systems for the case $G = SL(2)$, the most important for the initial group-theoretic problem. Towards this end, we

- prove several surjectivity theorems for the classical trace map over finite fields;
- introduce a new method based on the trace map and these theorems.

This allowed us not only to explain the mechanism of the proofs from the above cited papers but to obtain a method for producing more sequences of the same nature.

These arithmetic-geometric considerations led us to a new notion of residual periodicity of a dynamical system which reflects its local-global behaviour. This concept will hopefully yield new results in the arithmetic of dynamical systems on algebraic varieties. Here we present some primary examples and propose some conjectures.

To be more precise, let F_{r+s} be the free group with basis $x_1, \dots, x_s, u_1, \dots, u_r$, and let

$$\mathcal{W} = \left\{ \begin{array}{l} w_1(x_1, \dots, x_s, u_1, \dots, u_r), \\ \dots, \\ w_r(x_1, \dots, x_s, u_1, \dots, u_r). \end{array} \right\} \tag{1}$$

be an r-tuple of words in F_{r+s}. Thus for any group G we obtain a self-map:

$$D_{\mathcal{W}} : G^{r+s} \to G^{r+s}, \tag{2}$$

$$(g_1, \dots, g_s, v_1, \dots, v_r) \mapsto (g_1, \dots, g_s, w_1(g_1, \dots, g_s, v_1, \dots, v_r), \dots, w_r(g_1, \dots, g_s, v_1, \dots, v_r).$$

Choosing G to be a linear algebraic group defined over some field k, we thus find a polynomial self-map of the underlying affine variety G^{r+s} attached to $\mathcal{W}$.

A set $M \subset G^{r+s}$ is called invariant if $D_{\mathcal{W}}(M) \subset M$.

For our purposes it is important to introduce initial conditions and, for every group G, a so-called forbidden set. Let $\mathcal{J} = (f_1(x_1, \dots, x_s), \dots, f_r(x_1, \dots, x_s))$ be words in F_s. Then given G and $(g_1, \dots, g_s) \in G^s$ we have an iterative sequence of r-tuples of elements of G:

$$e_0 = (f_1(g_1, \dots, g_s), \dots, f_r(g_1, \dots, g_s)), \dots,$$

$$e_{n+1} = (w_1(g_1, \dots, g_s, e_n), \dots, w_r(g_1, \dots, g_s, e_n)), \dots$$

We are interested in finding $(g_1, \dots, g_s)$ such that the sequence $e_0, e_1, \dots$ has certain properties. To find such $(g_1, \dots, g_s)$, we add s extra "tautological" variables and obtain a self-map as in (2).

Then given $\mathcal{W}$, G and $\mathcal{J}$, we have an iterative sequence:

$$e'_0 = (g_1, \dots, g_s, f_1(g_1, \dots, g_s), \dots, f_r(g_1, \dots, g_s)), \dots,$$

$$e'_{n+1} = D_{\mathcal{W}}(e'_n), \dots$$

The forbidden set consists of the choice of an invariant set $I_G \subset G^{r+s}$ for every group G.

We call the triple $D = (\mathcal{W}, \mathcal{J}, I_G)$ a *verbal dynamical system*. We are interested in invariant sets disjoint from I_G.

Remark 1.1. It is sometimes convenient to modify this general setup as follows.

(i) It may happen that the r-tuple $\mathcal{W}$ depends on less than $r + s$ variables (say, of $x_1, \dots, x_s$ only $x_1, \dots, x_t$, $t < s$, show up in $\mathcal{W}$ whereas the rest of the x_i only appear in the initial conditions

$\mathcal{J}$). In such a case, we will restrict our dynamical system to G^{r+t} (in particular, the forbidden set is also chosen inside G^{r+t}). See Example 1.4 below.

(ii) One can fix an s-tuple $(g^\circ := (g_1^\circ, \dots, g_s^\circ) \in G^s$ and consider the corresponding "fibre" of our dynamical system $D^0_{\mathcal{W}}\colon G^r \to G^r$ defined by

$$D^0_{\mathcal{W}}((v_1, \dots, v_r)) = (w_1(g_1^\circ, \dots, g_s^\circ, v_1, \dots, v_r), \dots, w_r(g_1^\circ, \dots, g_s^\circ, v_1, \dots, v_r)).$$

In particular, for $r = 1$ we arrive at a self-map $G \to G$. This simplified system will be largely used in what follows.

Example 1.2. Take $s = 2$, $r = 1$ and consider a triple D_1 consisting of

$$\mathcal{W} = ([xux^{-1}, yuy^{-1}]),$$
$$\mathcal{J} = (x^{-2}y^{-1}x),$$
$$I_G = \{G \times G \times \{1\}\}.$$

The corresponding map is

$$D_{\mathcal{W}}(x, y, u) = (x, y, [xux^{-1}, yuy^{-1}]).$$

The associated iterative sequence is

$$e_0 = x^{-2}y^{-1}x, \quad e_1 = [x^{-1}y^{-1}, yx^{-2}y^{-1}xy^{-1}], \quad e_2 = [xe_1x^{-1}, ye_1y^{-1}], \dots$$

A key step in our characterization of finite solvable groups [BGGKPP1], [BGGKPP2] can now be reformulated as follows:

Theorem 1.3. *For $G = SL(2, q)$ the dynamical system D_1 has a fixed point outside I_G for every $q > 3$.*

A key step to the characterization obtained in [BWW] can be reformulated in a similar way:

Example 1.4. Take $s = 2$, $r = 1$, $\mathcal{W} = ([y^{-1}uy, u^{-1}])$, $\mathcal{J} = (x)$. As the variable x does not show up in $\mathcal{W}$ but only appears in $\mathcal{J}$ (and so $t = 1$), we proceed as in Remark 1.1(i) and consider the restricted system $G^2 \to G^2$, $(y, u) \mapsto (y, [y^{-1}uy, u^{-1}])$, with the forbidden set $I_G := \{G \times \{1\}\}$. Denote this system by D_2.

The associated iterative sequence is

$$e_0 = x, \quad e_1 = [y^{-1}xy, x^{-1}], \quad e_2 = [y^{-1}e_1y, e_1^{-1}], \dots$$

The main result of [BWW] can now be read off as follows:

Theorem 1.5. *For $G = SL(2, q)$ the dynamical system D_2 has a periodic point outside I_G for every $q > 3$.*

In the present paper we mostly restrict ourselves to considering the most important case $G = SL(2, k)$ (though in Section 4 we also consider the Suzuki groups).

In the case $G = SL(2, k)$ we introduce a new method based on classical results of Klein, Fricke, Vogt, Magnus from which it follows (see, e.g., [Pe2]) that there is a polynomial map $\psi\colon \mathbb{A}^N(k) \to \mathbb{A}^N(k)$ making the diagram

$$\begin{array}{ccc} G^{s+r} & \xrightarrow{D_{\mathcal{W}}} & G^{s+r} \\ {\scriptstyle\pi}\downarrow & & {\scriptstyle\pi}\downarrow \\ \mathbb{A}^N(k) & \xrightarrow{\psi} & \mathbb{A}^N(k) \end{array} \tag{3}$$

commutative. Here π is defined using the traces of products as in Theorem 3.1 below.

In the case $r = 1$, $t = 1$ the projection $\pi\colon SL(2, k)^2 \to \mathbb{A}^3(k)$ is defined as

$$\pi(x, y) = (\mathrm{tr}(x), \mathrm{tr}(xy), \mathrm{tr}(y)).$$

In the case $r = 1$, $s = 2$ the map $\pi\colon SL(2,k)^3 \to \mathbb{A}^7(k)$ is defined as

$$\pi(x,y,u) = (\mathrm{tr}(x), \mathrm{tr}(y), \mathrm{tr}(u), \mathrm{tr}(xy), \mathrm{tr}(xu), \mathrm{tr}(yu), \mathrm{tr}(xyu)),$$

and the image of π is contained in a hypersurface $Z \subset \mathbb{A}^7$ (see (13) below for an explicit equation of Z).

We prove the following surjectivity theorems (see Theorems 3.4 and 3.10 below).

Surjectivity Theorem 1. *For any point $\overline{a} = (s_0, u_0, t_0) \in \mathbb{A}^3(\mathbb{F}_q)$ the set $\pi^{-1}(\overline{a}) \subset SL(2,q)^2$ is nonempty.*

Surjectivity Theorem 2. *For any point $y \in Z(\mathbb{F}_q)$ the set $\pi^{-1}(y) \subset SL(2,q)^3$ is nonempty.*

These surjectivity theorems allow us to obtain sufficient conditions for the existence of fixed points of the reduced (modulo p) dynamical system, uniformly in p, and treat concrete examples arising from [BGGKPP1], [BGGKPP2], [BWW].

On the other hand, the above dynamical reinterpretation of our group-theoretic problem leads to some interesting "local-global" properties of dynamical systems on algebraic varieties. By an *AG dynamical system* (AG stands for arithmetic-geometric) we mean a triple $D = (X, V, \varphi)$, where

- either X is an algebraic variety defined over a global field K, $\varphi\colon X \to X$ is a dominant endomorphism and $V \subset X(K)$ is a subset invariant under φ;
- or X is an $\mathcal{O}$-scheme ($\mathcal{O}$ stands for the ring of integers in K), $\varphi\colon X \to X$ is dominant and $V \subset X(\mathcal{O})$ is a φ-invariant subset.

A periodic point is a fixed point of an iteration $\varphi^{(n)}$ of φ. Together with the system $D = (X, V, \varphi)$, we consider its reductions $D_p = (X_p, V_p, \varphi_p)$, where p ranges over all but finitely many places of K (see Section 6 for precise definitions). For each reduction, we consider the length ℓ_p of the shortest orbit C_p which does not intersect the "forbidden" set $V_p \subset X_p$. If such an orbit does not exist, we set $\ell_p = \infty$. We are interested in the distribution of ℓ_p's. More specifically, let $M \subset \mathbb{N}$ be the set of all primes p such that $\ell_p = \infty$. Let $N = \{\ell_p : p \notin M\}$.

- If M is infinite, we call the system **residually aperiodic**.
- If M is finite, we call the system **residually periodic**.
- If both M and N are finite, we call the system **strongly residually periodic**.

Precise definitions, examples and discussion of these notions are the subject of Section 6.

Remark 1.6. According to a theorem of Hrushovski [Hr], φ has a periodic point in $X(\overline{\mathbb{F}}_p) \setminus V(\overline{\mathbb{F}}_p)$ provided X is an affine $\mathbb{F}_p$-variety and V is a proper affine subset of X ($\overline{\mathbb{F}}_p$ stands for the algebraic closure of $\mathbb{F}_p$). In contrast, we are only interested in periodic points in $X(\mathbb{F}_p)$.

In this language our approach to the problem of characterization of finite solvable groups looks as follows. We consider word maps of groups $G = SL(2,q)$. For every word map $\varphi\colon G^m \to G$, $m = 2, 3$ (and an additional word $f\colon G^2 \to G$ in the case $m = 3$) we define a verbal dynamical system (see, e.g., Examples 1.2, 1.4). Regarding the group as an affine variety, we obtain from a verbal dynamical system an AG dynamical system on an affine $\mathbb{Z}$-scheme. (In Example 1.2 we have $X = SL(2) \times SL(2) \times SL(2)$, $V = SL(2) \times SL(2) \times \{1\}$, $\varphi(x,y,u) = (x, y, [xux^{-1}, yuy^{-1}])$, in Example 1.4 we have $X = SL(2) \times SL(2)$, $V = SL(2) \times \{1\}$, $\varphi(y,u) = (y, [y^{-1}uy, u^{-1}])$.) The word map is a "good" candidate if and only if that system is residually periodic. Using the trace map we simplify the AG system by including it into a commutative diagram

$$\begin{array}{ccc} X & \xrightarrow{\tilde{\varphi}} & X \\ {\scriptstyle\pi}\downarrow & & {\scriptstyle\pi}\downarrow \\ Y & \xrightarrow{\psi} & Y \end{array} \tag{4}$$

where π is a surjective projection, defined over $\mathbb{Z}$, and ψ is the trace map (see Subsections 3.1, 3.2 for more details). Moreover, the dynamical system $D' = (Y, \pi(V), \psi)$ has special geometric properties allowing us to find out when it is strongly residually periodic. Note that π is surjective, therefore if D' is strongly residually periodic then D is residually periodic.

It is an interesting question what arithmetic or geometric conditions can guarantee residual periodicity (or aperiodicity) of a given dynamical system. Certainly, if the forbidden set V is empty then the system is residually periodic.

The role of arithmetic may be demonstrated by the following example.

Example 1.7. Let a and b denote distinct integers, and let $H(x) = (x^2 - a)(x^2 - b)(x^2 - ab) + x$. The polynomial $H(x)$ defines a morphism $H\colon \mathbb{A}^1_{\mathbb{Z}} \to \mathbb{A}^1_{\mathbb{Z}}$.

For every p the reduced morphism H_p has fixed points. Indeed, if $p|a$ or $p|b$, we have $H_p(0) = 0$. If none of a and b is divisible by p, we can use the fact that the Legendre symbol is a multiplicative function and conclude that at least one of three numbers: a, b, ab, is a square modulo p. A square root of this number is then a fixed point of H_p, so we have $\ell_p = 1$.

On the other hand, the morphism $H\colon \mathbb{A}^1_{\mathbb{Z}} \to \mathbb{A}^1_{\mathbb{Z}}$ may have no periodic points. Indeed, according to [Na], the period of a rational point for a monic polynomial cannot exceed 2, and Magma computations show that for $a = 2$, $b = 3$ there is no rational solution to the equation $H(H(x)) - x = 0$.

This example shows that one of the reasons for residual periodicity may be the existence of periodic points defined over a splitting field. Polynomials of that kind were studied in [BB], [Br], [BBH], [So].

As to geometric conditions, the dynamical system under consideration may happen to be residually periodic because of the existence of invariant functions (say, when there is an "extra" coordinate on which φ acts trivially) as in the following simple example.

Example 1.8. Let $D = (X, V, \varphi)$, where $X = \mathbb{A}^2$, $V = \{(a, b) \in X : a = \pm 1 \text{ or } b = \pm 1 \text{ or } a = 0 \text{ or } b = 0\}$, and $\varphi(a, b) = (a^2 b, b)$. Consider the integral model $\mathcal{D} = (\mathcal{X}, \mathcal{V}, \Phi)$ where $\mathcal{X} = A^2_{\mathbb{Z}}$, $\mathcal{V} = \{(a, b) \in X(\mathbb{Z}) : a = \pm 1 \text{ or } b = \pm 1 \text{ or } a = 0 \text{ or } b = 0\}$ and $\Phi(a, b) = (a^2 b, b)$. We have $M = \{2, 3\}$. The variety of fixed points of Φ is a curve $C = \{(a, b) : ab = 1\}$, $C \bigcap \mathcal{V} = \{\pm(1, 1)\}$. Nevertheless, for any prime $p > 3$ we have $C_p \setminus V_p \neq \emptyset$, i.e. $\ell_p = 1$.

These examples show that there are at least two general reasons for a dynamical system to be strongly residually periodic. The first one is purely arithmetic as in Example 1.7. Our first observations show that even in the simplest cases of one-dimensional systems, arising questions are related to nontrivial arithmetical problems. In the case of elliptic curves, one of such problems has been solved by N. Jones by establishing a weakened version of the long-standing Koblitz's conjecture (see the appendix to the present paper).

The second one is of geometric nature as for the trace map above. This map has an invariant function which leads to the dimension jump for the variety of fixed points. Once we can prove that this variety W is absolutely irreducible (or at least contains an absolutely irreducible component), we can apply the Lang–Weil estimates [LW] to conclude that there exists a fixed point on the reduction W_q for q big enough. (Of course, if $\dim W = 1$, classical Weil's estimates (see, e.g., [FJ]) are quite enough.)

We believe that residually periodic dynamical systems is an object worthy of investigation. The following particular case seems to be especially interesting. Consider a $\mathbb{Z}$-scheme X, a dominant endomorphism φ of X, and define V as the union of all finite φ-orbits in $X(\mathbb{Z})$. Then V_p is the union of orbits of the reductions of all preperiodic points of φ. In simple words, this means that in this case we are interested in the distribution of the smallest periods of the maps φ_p not coming from preperiodic points of φ. To the best of our knowledge, such a classification of dynamical systems according to their "hidden" periodicity did not appear in the literature.

The structure of the paper is as follows.

Section 3 contains a general framework of our method for the most important case $G = PSL(2, q)$. The Suzuki groups are treated in Section 4. Applications to concrete sequences are contained in Section 5. Section 6 is completely devoted to the new notion of residually periodic dynamical systems. We give basic definitions, consider simple examples and state some conjectures. The appendix contains a theorem of N. Jones answering one the questions posed in Section 6.

2. Notation and preliminaries

Recall that in [BGGKPP1], [BGGKPP2], [BWW] there have been exhibited explicit families $\alpha_n(x, y)$, $\beta_n(x, y)$ of words in F_2 allowing one to characterize the class $\mathcal{S}$ of finite solvable groups in the class of all finite groups as follows:

A finite group G belongs to $\mathcal{S}$ if and only if there exists n such that G satisfies the identity $\gamma_n(x, y) := [\alpha_n(x, y), \beta_n(x, y)] \equiv 1$.

Here $[a, b] = aba^{-1}b^{-1}$ denotes the commutator.

As in the introduction, we produce these recurrence formulas using the dynamical viewpoint. We consider the dynamical systems D_1 and D_2 from Examples 1.2 and 1.4, respectively, and consider their fibres as in Remark 1.1(ii). This means that for any group G we introduce the maps $G \to G$: $\rho_{u,v}(w) := [uwu^{-1}, vwv^{-1}]$, $\sigma_u(w) := [u^{-1}wu, w^{-1}]$. Then the n-th term of the characterizing sequence can be written as the n-th iteration of the map ρ (resp. σ):

$$\gamma_n(x, y) = \rho_{x,y}^{(n)}(\gamma_0(x, y)) \tag{5}$$

(resp.

$$\gamma_n(x, y) = \sigma_y^{(n)}(\gamma_0(x, y))), \tag{6}$$

where $\gamma_0(x, y) = x^{-2}y^{-1}x$ (resp. $\gamma_0(x, y) = x$).

Suppose that S is a solvable group of derived length n. Then the recursive structure of the above formulas shows that $\gamma_n(x, y) \equiv 1$ in S. To establish the converse statement, it is enough to show that the identity $\gamma_n(x, y) \equiv 1$ does not hold in any finite minimal simple non-solvable group G. (That is precisely what was done in [BGGKPP1], [BGGKPP2], [BWW].)

To establish this fact in the case of sequences of type (6), it is enough to show that there exists $u = y_0 \in G$ such that the map σ_u has a (non-identity) **periodic** point, i.e. there exist a positive integer m and an element $1 \neq g \in G$ such that g can be written in the form $g = \gamma(x, y_0)$ and $\sigma_{y_0}^{(m)}(g) = \sigma_{y_0}(g)$. (For sequence (6), that is precisely what was done in [BWW].) It is important to note here that every point has a finite orbit (i.e. is preperiodic in the sense of [Si1]) but *a priori* it can happen that all these orbits contain identity, which being fixed is the only periodic point. We need an orbit that never hits the identity, and therefore contains another periodic point. This explains our choice of the forbidden set in Examples 1.2 and 1.4.

Let us recall the list of minimal simple non-solvable groups [Th]:

(1) $G = PSL(2, p)$, $p = 5$ or $p \equiv \pm 2 \pmod 5$, $p \neq 3$,
(2) $G = PSL(2, 2^p)$,
(3) $G = PSL(2, 3^p)$, p is an odd prime,
(4) $G = Sz(2^p)$, p is an odd prime,
(5) $G = PSL(3, 3)$.

Here Sz stands for the Suzuki group (twisted form of B_2, see, e.g., [HB] for details).

To obtain a characterization of finite solvable groups, we wish to find a word $\varphi \in F_2(x, y)$ with the following properties:

(i) for any finite solvable group S there exists an integer n such that for every $y \in S$ the map $\varphi_y^{(n)} \colon S \to S$ is the identity map (here $\varphi_y(x) := \varphi(x, y)$);

(ii) for each finite simple non-solvable group G from the above list, there exists $y \in G$ such that the self-map $\varphi_y \colon G \to G$ has a non-identity periodic point. For the $PSL(2)$ case, this fits into the

approach described in Section 1: we consider the dynamical system $(PSL(2,\mathbb{Z}),\{1\},\varphi_y)$ and all its reductions. (Note that in our context, the difference between SL and PSL is negligible, see Remark 3.22 below.)

In order to satisfy condition (i), one has to impose some restrictions on φ. We shall discuss this matter in Section 6.

In the sequel, we shall consider two separate cases: $G = PSL(2,q)$ and $G = Sz(q)$ (the case of the single group $G = PSL(3,3)$ is usually easy to handle). In each case we will show that the corresponding dynamical system D gives rise to a dynamical system D' in the space of traces (the trace map) as in diagram (4). The trace map has special geometry: the set of its fixed points (or of periodic points of bounded period) has positive dimension. This allows us to formulate a geometric sufficient condition on φ in order to get a periodic point in every reduction. (See Section 6 where we dare formulate some general conjectures.)

Further on we denote by $\mathbb{A}^n_{x_1,\dots,x_n}$ the affine space with coordinates $x_1,\dots,x_n$.

For brevity, we denote $\widetilde{G} = SL(2,q)$.

We will repeatedly use expressions of the form "a rational curve with n punctures" (even if our curve lies in an affine space) referring to an open subset of a projective curve of genus zero whose complement consists of n points (e.g., the curve $xy = 1$ in the affine plane will be referred to as a rational curve with two punctures).

3. Case $G = PSL(2,q)$

In this section we show how every word map gives rise to a dynamical system. Then we prove that this dynamical system may be included into a commutative diagram of type (4) (namely, diagrams (8) and (14) below). The idea is that it is sufficient to look for periodic points of the trace map ψ. Indeed, if a point a is ψ-periodic, then all the points in the fibre over a are φ-periodic. The problem is to show that this fibre is not empty. We first show how to construct the trace map, then we show that the projection is a surjective morphism for every reduction (Theorems 3.4 and 3.10). Specific geometry of the trace map allows us to give sufficient conditions for the correewponding dynamical system to be residually periodic (Theorems 3.6 and 3.21).

Our method is based on the following classical fact ([Vo], [Fr], [FK], [Ma1]) cited here from the paper [Ho] (see also [Ma2], [Go] for a nice modern exposition of these results).

Theorem 3.1. *Let $F = \langle a_1,\dots,a_n\rangle$ denote the free group on n generators. Let us embed F into $SL(2,\mathbb{Z})$ and denote by* tr *the trace character. If u is an arbitrary element of F, then the character of u can be expressed as a polynomial*

$$\mathrm{tr}(u) = P(t_1,\dots,t_n,t_{12},\dots,t_{12\dots n})$$

with integer coefficients in the 2^n-1 characters $t_{i_1 i_2\dots i_\nu} = \mathrm{tr}(a_{i_1}a_{i_2}\dots a_{i_\nu})$, $1\le\nu\le n$, $1\le i_1 < i_2 < \dots < i_\nu \le n$. □

Note that the theorem remains true for the group $\widetilde{G} = SL(2,q)$ (and, more generally, for $SL(2,R)$ where R is any commutative ring, see [CMS]).

We shall use this theorem in two different situations: for maps arising from formulas of type (6), called two-variable maps, and for those arising from formulas of type (5), called three-variable maps. These situations will be described in the next two subsections respectively.

3.1. Two-variable maps.

In this section we focus on the underlying affine algebraic variety of the algebraic group $\widetilde{G}$. Consider a morphism $\varphi\colon \widetilde{G}\times\widetilde{G}\to\widetilde{G}$ satisfying the property (needed for descending to $G = PSL(2)$):

$$\varphi(\pm x,\pm y) = \pm\varphi(x,y).$$

For example, any word map provides such a morphism. Namely, for any $x, y \in \widetilde{G}$ denote $s = \mathrm{tr}(x)$, $t = \mathrm{tr}(y)$, and $u = \mathrm{tr}(xy)$, and define a morphism $\pi\colon \widetilde{G} \times \widetilde{G} \to \mathbb{A}^3_{s,u,t}$ by

$$\pi(x, y) := (s, u, t).$$

Then in view of Theorem 3.1 there exists a map $\psi\colon \mathbb{A}^3_{s,u,t} \to \mathbb{A}^3_{s,u,t}$ such that

$$\psi(\pi(x, y)) = \pi(\varphi(x, y), y). \tag{7}$$

This map is called a "trace map" and is widely used (see, e.g., [Pe2]).

Define $\tilde{\varphi} = (\varphi, id)\colon \widetilde{G} \times \widetilde{G} \to \widetilde{G} \times \widetilde{G}$ by $\tilde{\varphi}(x, y) = (\varphi(x, y), y)$. Then the following diagram commutes:

$$\begin{array}{ccc} \widetilde{G} \times \widetilde{G} & \xrightarrow{\tilde{\varphi}} & \widetilde{G} \times \widetilde{G} \\ {\scriptstyle\pi}\downarrow & & {\scriptstyle\pi}\downarrow \\ \mathbb{A}^3_{s,u,t} & \xrightarrow{\psi} & \mathbb{A}^3_{s,u,t} \end{array} \tag{8}$$

Here $\psi(s, u, t) := (f_1(s, u, t), f_2(s, u, t), t)$, where $f_1(s, u, t) = \mathrm{tr}(\varphi(x, y))$, $f_2(s, u, t) = \mathrm{tr}(\varphi(x, y)y)$.

Lemma 3.2. *For any word map $\varphi(x, y)$ the variety*

$$\Phi : \{f_1(s, u, t) = s, f_2(s, u, t) = u\} \subset \mathbb{A}^3_{s,u,t}$$

of fixed points of ψ has positive dimension.

Proof. Since the variety Φ is defined by two equations in $\mathbb{A}^3_{s,u,t}$, it is sufficient to show that it is not empty. But for any word $\omega(x, y)$ we have: $\omega(1, 1) = 1$, thus $\psi(2, 2, 2) = (2, 2, 2)$, hence $\Phi \neq \emptyset$. □

Lemma 3.3. *Let $Q = (s_0, u_0, t_0)$ be a fixed point of ψ defined over $\mathbb{F}_q$. Let $(x, y) \in \pi^{-1}(Q)$. Then $(\varphi(x, y), y) \in \pi^{-1}(Q)$ as well.*

Proof. Indeed, (7) gives $\pi(\varphi(x, y), y) = \psi(Q) = Q$. □

Theorem 3.4. *For every $\mathbb{F}_q$-rational point $Q = (s_0, u_0, t_0) \in \mathbb{A}^3_{s,u,t}$ the fibre $H = \pi^{-1}(Q)$ has an $\mathbb{F}_q$-rational point.*

Proof. We will look for an element of H among pairs of matrices of the form

$$\left(\begin{pmatrix} 0 & 1 \\ -1 & s_0 \end{pmatrix}, \begin{pmatrix} a & b \\ c & -a + t_0 \end{pmatrix} \right). \tag{9}$$

To lie in H, the entries of these matrices must satisfy the equations

$$a(-a + t_0) - bc = 1, \quad c - b + s_0(-a + t_0) = u_0.$$

On eliminating b, we arrive at the following equation in a and c:

$$a^2 + c^2 - s_0 ac - t_0 a + (s_0 t_0 - u_0)c + 1 = 0, \tag{10}$$

which has a solution for every q. Of course, this can be proved using the Chevalley–Warning theorem, but for the reader's convenience we present here an elementary proof.

Case 1. q is odd.

The discriminant D of the quadratic part of the left-hand side of (10) equals $s_0^2 - 4$. If $D = 0$, i.e. $s_0 = \pm 2$, we exhibit an explicit point in H:

$$\left(\begin{pmatrix} \pm 1 & u_0 \mp t_0 \\ 0 & \pm 1 \end{pmatrix}, \begin{pmatrix} 1 & t_0 - 2 \\ 1 & t_0 - 1 \end{pmatrix} \right), \tag{11}$$

so we may assume $D \neq 0$. First, by a linear change of variables over $\mathbb{F}_q$, let us bring (10) to the form

$$\tilde{a}^2 + \varepsilon \tilde{c}^2 = r.$$

If r is a square, $r = v^2$, we can put $\tilde{a} = v$, $\tilde{c} = 0$, so we may assume that r is not a square. If ε is not a square, then r/ε is a square, $r/\varepsilon = v^2$, and we can put $\tilde{a} = 0$, $\tilde{c} = v$, so we may assume ε is a square, $\varepsilon = v^2$. In $\mathbb{F}_q$ there are $(q+1)/2$ squares and $(q-1)/2$ nonsquares, thus among $(q+1)/2$ elements $r - \tilde{a}^2$, when $\tilde{a}$ ranges over $\mathbb{F}_q$, there is a square w^2. We then put $\tilde{c} = w/v$.

Case 2. q **is even**.

If $s_0 = 0$, then we get a point in H from (11), so we may assume $s_0 \neq 0$. Then on putting $\tilde{a} = a + (s_0t_0 + u_0)/s_0$, $\tilde{c} = c + t_0/s_0$, we bring (10) to the form

$$\tilde{a}^2 + \tilde{c}^2 + s_0\tilde{a}\tilde{c} = r.$$

As every element of $\mathbb{F}_q$ is a square, we have $r = v^2$ and we can put $\tilde{a} = v$, $\tilde{c} = 0$. □

Corollary 3.5. *Consider the following "conjugation" equivalence relation $\sim$ on $SL(2,\mathbb{F}_q)^2$:*

$$(x,y) \sim (x',y') \text{ iff } \exists g \in SL(2,\overline{\mathbb{F}}_q) \mid x' = gxg^{-1},\ y' = gyg^{-1}.$$

Then every absolutely irreducible component of the set of conjugacy classes of $\tilde{\varphi}$-periodic points is positive dimensional.

Proof. Indeed $(SL(2,\mathbb{F}_q)^2 \setminus V(\mathbb{F}_q))/_\sim$ can be identified with $\mathbb{F}_q^3$. The corollary is valid, because the periodic set of the trace map is positive dimensional. □

We can now obtain a sufficient condition for the existence of periodic points. Consider the maps $\varphi\colon \widetilde{G} \times \widetilde{G} \to \widetilde{G}$ and $\psi\colon \mathbb{A}^3_{s,u,t} \to \mathbb{A}^3_{s,u,t}$ as in diagram (8), and denote by $\Phi \subset \mathbb{A}^3_{s,u,t}$ the variety of fixed points of ψ. As in Section 2, for a fixed y denote by $\varphi_y\colon \widetilde{G} \to \widetilde{G}$ the map $x \mapsto \varphi(x,y)$.

Note that Φ contains a line

$$L_1 = \{s = 2, u = t\}.$$

Since Φ is a complete intersection, all its irreducible components have dimension at least one.

Theorem 3.6. *Write $\Phi = \bigcup_{i=1}^{k} W_i \cup L_1$, where W_i are irreducible $\mathbb{F}_q$-components of Φ. Suppose q is big enough. If at least one of W_i's is absolutely irreducible, then there exists a pair $(x,y) \in G \times G$ such that $x \neq 1, y \neq 1$ and x is a periodic point of φ_y.*

Proof. Let W_i be an absolutely irreducible component of W, $W \neq L_1$. By the Lang–Weil theorem [LW], there is a point $Q = (s_0,u_0,t_0) \neq (\pm 2, t, \pm t) \in W_i(\mathbb{F}_q)$. According to Theorem 3.4, we have $H_Q(\mathbb{F}_q) \neq \emptyset$, where $H_Q = \pi^{-1}(Q)$. It follows that there exists a pair $(x,y) \in \widetilde{G} \times \widetilde{G}$ such that $s_0 = \operatorname{tr}(x)$, $u_0 = \operatorname{tr}(xy)$, $t_0 = \operatorname{tr}(y)$. By Lemma 3.3, $(\varphi_y(x), y) \in H_Q(\mathbb{F}_q)$ as well. Since the set $H_Q(\mathbb{F}_q)$ is finite, there are numbers $n < m \in \mathbb{N}$ such that $\varphi_y^{(m)}(x) = \varphi_y^{(n)}(x)$. Thus, $\tilde{x} = \varphi_y^{(n)}(x)$ is a periodic point of φ_y. Moreover, the image of $\tilde{x}$ in $G = PSL(2,q)$ is non-identity since $Q = (s_0,u_0,t_0) \neq (\pm 2, t, \pm t)$. □

Remark 3.7. If there is a component $W_i \subset \Phi$ defined over $\mathbb{Z}$ and irreducible over $\overline{\mathbb{Q}}$, then, by [Gr, Theorem IV, 9, 7.7(i)], the assumptions of the theorem are satisfied for any prime p big enough.

Remark 3.8. Suppose $q = p > 3$ is a prime number. Note that all the maps in diagram (8) are defined over $\mathbb{Z}$, and it can thus be viewed as the special fibre at p of the following diagram of morphisms of $\mathbb{Z}$-schemes (denoted by the same letters):

$$\begin{array}{ccc} \mathcal{G} \times \mathcal{G} & \xrightarrow{\tilde{\varphi}} & \mathcal{G} \times \mathcal{G} \\ \pi\downarrow & & \pi\downarrow \\ \mathbb{A}^3_{\mathbb{Z}} & \xrightarrow{\psi} & \mathbb{A}^3_{\mathbb{Z}} \end{array} \tag{12}$$

where $\mathcal{G} = SL(2,\mathbb{Z})$.

3.2. **Three-variable maps.** Let here $\widetilde{G}$ denote $SL(2,K)$ where K is an arbitrary field. Consider a morphism $\varphi\colon \widetilde{G}\times\widetilde{G}\times\widetilde{G}\to\widetilde{G}$ such that

$$\varphi(\pm x,\pm u,\pm y)=\pm\varphi(x,u,y).$$

The modified map $\tilde{\varphi}\colon \widetilde{G}\times\widetilde{G}\times\widetilde{G}\to\widetilde{G}\times\widetilde{G}\times\widetilde{G}$ is defined by $\tilde{\varphi}(x,u,y)=(x,\varphi(x,u,y),y)$.

As above, we consider a representation ρ of the free group F_3 in $SL(2,\mathbb{Z})$ and assume that φ is defined by a word $w=w(x,u,y)$. The trace of $\rho(w)$ can be expressed as a polynomial in 7 variables $a_1=\mathrm{tr}(x)$, $a_2=\mathrm{tr}(y)$, $a_3=\mathrm{tr}(u)$, $a_{12}=\mathrm{tr}(xy)$, $a_{13}=\mathrm{tr}(xu)$, $a_{23}=\mathrm{tr}(yu)$, $a_{123}=\mathrm{tr}(xyu)$. These variables are dependent (see, e.g., [Ma1] or formulas (2.3)–(2.5) in [Ho]):

$$\begin{aligned}&a_{123}^2-a_{123}(a_{12}a_3+a_{13}a_2+a_{23}a_1-a_1a_2a_3)\\&\quad+(a_1^2+a_2^2+a_3^2+a_{12}^2+a_{13}^2+a_{23}^2-a_1a_2a_{12}-a_1a_3a_{13}-a_2a_3a_{23}+a_{12}a_{13}a_{23}-4)=0.\end{aligned}\tag{13}$$

Let $\overline{a}=(a_1,a_2,a_3,a_{12},a_{13},a_{23},a_{123})\in\mathbb{A}^7$, let $Z\subset\mathbb{A}^7$ be an absolutely irreducible set defined by (13). Let $\pi(x,u,y)=\overline{a}\in Z$ be the trace projection. Then the following diagram is commutative:

$$\begin{array}{ccc}\widetilde{G}\times\widetilde{G}\times\widetilde{G} & \xrightarrow{\tilde{\varphi}} & \widetilde{G}\times\widetilde{G}\times\widetilde{G}\\ \pi\big\downarrow & & \pi\big\downarrow\\ Z(K) & \xrightarrow{\psi} & Z(K)\end{array}\tag{14}$$

where $\psi(\overline{a})=(a_1,a_2,l_1(\overline{a}),a_{12},l_2(\overline{a}),l_3(\overline{a}),l_4(\overline{a}))$,

$$l_1=\mathrm{tr}(\varphi(x,u,y)),\ \ l_2=\mathrm{tr}(\varphi(x,u,y)x),$$
$$l_3=\mathrm{tr}(\varphi(x,u,y)y),\ \ l_4=\mathrm{tr}(\varphi(x,u,y)xy).$$

The variety $F(\varphi)\subset Z$ of fixed points of ψ is defined by the equations

$$l_1(\overline{a})=a_3,\ \ l_2(\overline{a})=a_{13},\ \ l_3(\overline{a})=a_{23},\ \ l_4(\overline{a})=a_{123},$$

and, since it is nonempty, its dimension is at least 3.

Let us now consider diagram (14) more carefully.

Lemma 3.9. *Let F be any algebraically closed field. Then the set Z is an irreducible hypersurface over F.*

Proof. Assume the contrary. Let p denote the natural projection of $\mathbb{A}^7$ to $\mathbb{A}^6$, forgetting the coordinate a_{123}. Let $L\subset\mathbb{A}^6$ be an irreducible curve not contained in the branch locus of the restriction of p to Z. Then the set $p^{-1}(L)\bigcap Z$ is reducible.

Case 1. *$char(F)\neq 2$.*

Let $c\neq\pm 2$. Consider the curves $L=\{a_1=a_2=a_{13}=a_{23}=0, a_{12}=c\}\subset\mathbb{A}^6$ and $M=p^{-1}(L)=\{a_1=a_2=a_{13}=a_{23}=0, a_{12}=c\}\subset\mathbb{A}^7$. Then from (13) it follows that $M'=Z\bigcap M$ is defined by the following equations:

$$(a_{123}-a_3c/2)^2-(c^2-4)(a_3^2-4)/4=0,\ a_1=a_2=a_{13}=a_{23}=0,\ a_{12}=c.$$

Therefore M' is a branched double cover of L, hence it is irreducible. Contradiction.

Case 2. *$char(F)=2$.* We now consider the curve $L=\{a_1=a_2=a_{13}=a_{23}=0, a_{12}=a_3+1\}\subset\mathbb{A}^6$. In the notation of Case 1, M' is defined by the equations

$$a_{123}^2-a_3(a_3+1)a_{123}+1=0,\ a_1=a_2=a_{13}=a_{23}=0,\ a_{12}=a_3+1.$$

Thus it is irreducible. Contradiction.

Hence Z is irreducible. □

Theorem 3.10. *Let $Z\subset\mathbb{A}^7_{a_1,a_2,a_3,a_{12},a_{13},a_{23},a_{123}}$ be defined by equation* (13)*. Then for all q the map $\pi\colon SL(2,q)\times SL(2,q)\times SL(2,q)\to Z(\mathbb{F}_q)$ is surjective.*

Proof. The result will follow from identities between certain polynomials in the polynomial ring

$$R := \mathbb{Z}[x_1, x_2, x_3, x_{12}, x_{13}, x_{23}, x_{123}, \alpha_1, \gamma_1, \alpha_2, \gamma_2].$$

Denote

$$\begin{aligned} L :=& x_{123}^2 - x_{123}(x_{12}x_3 + x_{13}x_2 + x_{23} + x_1 - x_1x_2x_3) \\ &+ x_1^2 + x_2^2 + x_3^2 + x_{12}^2 + x_{13}^2 + x_{23}^2 - x_1x_2x_{12} - x_1x_3x_{13} - x_2x_3x_{23} + x_{12}x_{13}x_{23} - 4, \\ L_{12} :=& x_1^2 + x_2^2 + x_{12}^2 - x_1x_2x_{12} - 4, \quad L_{13} := x_1^2 + x_3^2 + x_{13}^2 - x_1x_3x_{13} - 4, \\ & L_{23} := x_2^2 + x_3^2 + x_{23}^2 - x_2x_3x_{23} - 4 \end{aligned} \tag{15}$$

(all viewed as elements of R).

We start with the following lemma (skipping an elementary proof).

Lemma 3.11. *Let K be a finite field, and let r, s, t, $a \in K$ be such that the equation in x, y*

$$x^2 + y^2 + rxy + sx + ty = a$$

is not solvable in K. Then the characteristic of K is 2 and $r = 0$, $s = t$ hold. □

We now define two more polynomials in the ring R (the reason will become clear later on):

$$D_1 := -\alpha_1^2 + \alpha_1\gamma_1x_3 + \alpha_1x_1 - \gamma_1^2 - \gamma_1x_1x_3 + \gamma_1x_{13} - 1,$$
$$D_2 := -\alpha_2^2 + \alpha_2\gamma_2x_3 + \alpha_2x_2 - \gamma_2^2 - \gamma_2x_2x_3 + \gamma_2x_{23} - 1.$$

Our argument will also need the following two by two matrix over R:

$$A = \begin{pmatrix} 2\alpha_2 - \gamma_2x_3 - x_2 & -\alpha_2x_3 + 2\gamma_2 + x_2x_3 - x_{23} \\ \alpha_2x_3 - 2\gamma_2 - x_2x_3 + x_{23} & -\alpha_2x_3^2 + 2\alpha_2 + \gamma_2x_3 + x_2x_3^2 - x_2 - x_3x_{23} \end{pmatrix}. \tag{16}$$

Define further $\tilde{A}$ to be the adjoint matrix of A, that is $\tilde{A}$ is A with the diagonal entries permuted and the off-diagonal entries multiplied by -1. The product $\tilde{A}A$ is the scalar matrix corresponding to the determinant of A. We further consider the vector

$$b := \begin{pmatrix} \alpha_2x_1 - \gamma_2x_1x_3 + \gamma_2x_{13} - x_1x_2 + x_{12} \\ \alpha_2x_{13} - \gamma_2x_1 - x_2x_{13} + x_{123} \end{pmatrix} \in R^2$$

and define $r, s \in R$ by

$$\begin{pmatrix} r \\ s \end{pmatrix} := \tilde{A}b.$$

Multiply now D_1 by L_{23}^2 and replace $y_1 := L_{23}^2\alpha_1$, $y_2 := L_{23}^2\gamma_1$, obtaining the polynomial

$$F(y_1, y_2) := -y_1^2 + y_1y_2x_3 + y_1L_{23}x_1 - y_2^2 - y_2L_{23}x_1x_3 + y_2L_{23}x_{13} - L_{23}^2$$

in the variables y_1, y_2.

We need one more lemma.

Lemma 3.12. *Let $\mathfrak{D}_2$ be the ideal of R generated by D_2 and $\mathfrak{D}$ the ideal generated by D_2 and L. Then the following hold:*

(i) $\det(A) - L_{23}$ *is in* $\mathfrak{D}_2$*;*
(ii) $F(r, s)$ *is in* $\mathfrak{D}$*.*

The proof of this lemma amounts to certain simple computations which are best done using a computer algebra system. The first item follows for example from the identity:

$$\det(A) - L_{23} = (x_3^2 - 4)D_2.$$

For the second item, the formula is more complicated. We skip the details. □

We can now go over to the proof of the theorem.

Let K be any field. Let $x = (x_1, x_2, x_3, x_{12}, x_{13}, x_{23}, x_{123}) \in Z(K)$. As we are working with traces and are thus allowed to make simultaneous conjugation, we start our search of solutions to

$\pi(B_1, B_2, B_3) = x$ by considering the following triples of two by two matrices over the polynomial ring $K[\alpha_1, \gamma_1, \alpha_2, \gamma_2]$:

$$B_1 = \begin{pmatrix} \alpha_1 & -\alpha_1 x_3 + \gamma_1 + x_1 x_3 - x_{13} \\ \alpha_1 & x_1 - \alpha_1 \end{pmatrix}, \qquad B_2 = \begin{pmatrix} \alpha_2 & -\alpha_2 x_3 + \gamma_2 + x_2 x_3 - x_{23} \\ \gamma_2 & x_2 - \alpha_2 \end{pmatrix}, \tag{17}$$

$$B_3 = \begin{pmatrix} 0 & 1 \\ -1 & x_3 \end{pmatrix}. \tag{18}$$

The condition that B_1, B_2, B_3 have determinant 1 and satisfy $\pi(B_1, B_2, B_3) = x$ is equivalent to the four equations:

$$D_1 = -\alpha_1^2 + \alpha_1\gamma_1 x_3 + \alpha_1 x_1 - \gamma_1^2 - \gamma_1 x_1 x_3 + \gamma_1 x_{13} - 1 = 0, \tag{19}$$

$$D_2 = -\alpha_2^2 + \alpha_2\gamma_2 x_3 + \alpha_2 x_2 - \gamma_2^2 - \gamma_2 x_2 x_3 + \gamma_2 x_{23} - 1 = 0, \tag{20}$$

$$\begin{aligned} \alpha_1(2\alpha_2 - \gamma_2 x_3 - x_2) + \gamma_1(-\alpha_2 x_3 + 2\gamma_2 + x_2 x_3 - x_{23}) \\ - \alpha_2 x_1 + \gamma_2 x_1 x_3 - \gamma_2 x_{13} + x_1 x_2 - x_{12} = 0, \end{aligned} \tag{21}$$

$$\begin{aligned} \alpha_1(\alpha_2 x_3 - 2\gamma_2 - x_2 x_3 + x_{23}) + \gamma_1(-\alpha_2 x_3^2 + 2\alpha_2 + \gamma_2 x_3 + x_2 x_3^2 - x_2 - x_3 x_{23}) \\ - \alpha_2 x_{13} + \gamma_2 x_1 + x_2 x_{13} - x_{123} = 0. \end{aligned} \tag{22}$$

Notice that the first equation is quadratic in α_1, γ_1 only and the second is quadratic in α_2, γ_2 only. The third and fourth equations are written as a linear system in α_1, γ_1. Defining the vectors

$$y := \begin{pmatrix} \alpha_1 \\ \gamma_1 \end{pmatrix}, \qquad b := \begin{pmatrix} \alpha_2 x_1 - \gamma_2 x_1 x_3 + \gamma_2 x_{13} - x_1 x_2 + x_{12} \\ \alpha_2 x_{13} - \gamma_2 x_1 - x_2 x_{13} + x_{123} \end{pmatrix},$$

the third and fourth of the above equations can be schematically written as

$$Ay = b$$

with the matrix A defined in (16) evaluated at our point x.

We now assume that K is a finite field. We shall now write $L_{23}(x)$ for the polynomial L_{23} defined above evaluated at our point $x \in Z(K)$, that is $L_{23}(x) = x_2^2 + x_3^2 + x_{23}^2 - x_2 x_3 x_{23} - 4$. We use a similar notation for all the other polynomials.

Case 1: At least one of the values $L_{12}(x)$, $L_{13}(x)$, $L_{23}(x)$ is nonzero. Assume, say, $L_{23}(x) \neq 0$ (the other cases are similar).

First we show that (20), viewed as an equation in the indeterminates α_2, γ_2, has a solution. Assume the contrary. Then by Lemma 3.11 we conclude that the characteristic of K is two, $x_3 = 0$ and $x_2 = x_{23}$. This contradicts the assumption $L_{23} \neq 0$.

We shall now fix a solution $(\alpha_2, \gamma_2) \in K^2$ of equation (20) and put these into the above matrix A getting a two by two matrix over K. Similarly we get a vector b in K^2. By Lemma 3.12 we find

$$\det(A) = L_{23}(x) \neq 0$$

which is guaranteed by our assumption. We now define $(\alpha_1, \gamma_1) \in K^2$ by

$$\begin{pmatrix} \alpha_1 \\ \gamma_1 \end{pmatrix} := A^{-1} b.$$

By Lemma 3.12(ii), we have found three matrices $B_1,\ B_2,\ B_3 \in SL(2, K)$ satisfying $\pi(B_1, B_2, B_3) = x$.

If now $L_{23}(x) = 0$, we have either $L_{12}(x) \neq 0$ or $L_{13}(x) \neq 0$. These cases are treated in a similar way. □

Remark 3.13. The above proof remains true if K is any quadratically closed field (cf. also [Pe1]).

Case 2: $L_{12}(x) = L_{13}(x) = L_{23}(x) = 0$.

Loosely speaking, our strategy in this case is to use automorphisms of the free group F_3 to get from x another point of $Z(K)$ such that not all three values of L_{ij} vanish at that point, and then use the result of Case 1. Let us make this more precise.

We start with an obvious lemma.

Lemma 3.14. *Let $n \geq 2$, let F_n denote the free group on n generators $X_1, \ldots, X_n$, and let G^n be the product of n copies of a group G. The map*

$$\mathrm{Aut}(F_n) \to \mathrm{Sym}(G^n), \quad \varphi \mapsto \hat{\varphi},$$

defined by

$$\hat{\varphi}(T) = (\varphi(X_1)_T, \ldots, \varphi(X_n)_T),$$

is a group homomorphism.

Here T is an n-tuple of elements of G and $\varphi(X_i)_T$ is the element of G obtained by substitution of the elements of T instead of the X_i's appearing in the expression of $\varphi(X_i)$ in the basis $\{X_1, \ldots, X_n\}$. □

The following constructions are described in [Ho] (see also [Ma1], [Pe2]), sometimes with details omitted. For the reader's convenience and sake of completeness we now focus on the case $n = 3$ giving some more details. Fix a basis $\{X, Y, Z\}$ of F_3.

Definition 3.15. For every $\varphi \in \mathrm{Aut}(F_3)$ define a map $F_\varphi \colon \mathbb{A}^7 \to \mathbb{A}^7$ by the formula

$$F_\varphi(u) := (P_{\varphi(X)}(u), P_{\varphi(Y)}(u), \ldots, P_{\varphi(XYZ)}(u)),$$

where P_w is the integer polynomial in 7 variables corresponding to the word w (cf. Theorem 3.1).

Lemma 3.16. *For every $\varphi \in \mathrm{Aut}(F_3)$ and every $T \in SL(2,K)^3$ we have*

$$\pi(\hat{\varphi}(T)) = F_\varphi(\pi(T)).$$

Proof. Obvious. □

Lemma 3.17. *For every $\varphi \in \mathrm{Aut}(F_3)$ and every field K we have $F_\varphi(Z(K)) \subseteq Z(K)$.*

Proof. We first prove that $F_\varphi(Z(\overline{K})) \subseteq Z(\overline{K})$, where $\overline{K}$ is an algebraic closure of K. From this the needed inclusion will follow as soon as F_φ is defined over K. In Case 1 we have proven that the map π is surjective onto an open subset

$$U(\overline{K}) = \{L_{12} \neq 0, L_{13} \neq 0, L_{23} \neq 0\} \subseteq Z(\overline{K}),$$

since the proof was valid for any algebraically closed field (see Remark 3.13).

Let $u \in U(\overline{K}), u = \pi(T), T \in SL(2,\overline{K})^3$. Then $F_\varphi(u) = F_\varphi(\pi(T)) = \pi(\hat{\varphi}(T)) \in Z(\overline{K})$. Hence, $F_\varphi(U(\overline{K})) \subseteq Z(\overline{K})$. Since U is open in Z and Z is irreducible, the same inclusion is valid for Z. Since F_φ is defined over $\mathbb{Z}$, the inclusion for K-points follows as well. □

Lemma 3.18. (i) $F_{\mathrm{id}} = \mathrm{id}$;

(ii) *For every $\varphi, \psi \in \mathrm{Aut}(F_3)$ and every $u \in Z(K)$ we have*

$$F_{\varphi\circ\psi}(u) = F_\varphi \circ F_\psi(u).$$

Proof. The first item is obvious, so let us prove the second one. Once again, similarly to Lemma 3.17, it is sufficient to prove it over an open subset U considered in Lemma 3.17, and over the algebraically closed field $\overline{K}$).

Let us take $u \in U(\overline{K}), u = \pi(T), T \in SL(2,\overline{K})^3$. Using Lemmas 3.14 and 3.16, we get

$$F_{\varphi\circ\psi}(u) = \pi(\widehat{\varphi \circ \psi}(T)) = \pi(\hat{\varphi} \circ \hat{\psi}(T)),$$

$$F_\varphi \circ F_\psi(u) = F_\varphi(\pi(\hat{\psi}(T)) = \pi(\hat{\varphi}(\hat{\psi}(T))),$$

so the needed equality is proved. □

Corollary 3.19. *The correspondence $\varphi \mapsto F_\varphi$ defines a group homomorphism $\mathrm{Aut}(F_3) \to \mathrm{Aut}(Z)$ where $\mathrm{Aut}(Z)$ is the group of $\mathbb{Z}$-defined polynomial automorphisms of the variety Z.* □

We can now go over to the proof of the theorem in Case 2.

Let, as above, $x \in Z(K)$ be such that $L_{12}(x) = L_{13}(x) = L_{23}(x) = 0$.

Case 2a. Let first assume that there exists $\varphi \in \mathrm{Aut}(F_3)$ such that $u := F_\varphi(x)$ is such that not all three values $L_{12}(u)$, $L_{13}(u)$, $L_{23}(u)$ are zero. By Case 1, there exists $T \in SL(2,K)^3$ such that $\pi(T) = u$. Define $T' := \hat{\varphi}^{-1}(T)$. By Lemma 3.16 and Corollary 3.19, we have $\pi(T') = F_{\varphi^{-1}}(\pi(T)) = F_{\varphi^{-1}}(u) = F_\varphi^{-1}(u) = x$, and we are done.

Case 2b. Assume that there is no such φ as in Case 2a.

Denote by L_{ij}^φ (where i, j stand for distinct numbers from the set $\{1,2,3\}$) the polynomials in 7 variables obtained after applying F_φ to L_{ij}. The needed contradiction immediately follows from the following proposition.

Proposition 3.20. *Denote the automorphisms of F_3 sending the basis $\{X,Y,Z\}$ to the bases $\{XY,Y,Z\}$, $\{X,YZ,Z\}$, $\{X,Y,XZ\}$, $\{XY^{-1},Y,Z\}$, $\{X,Y,YZ\}$, $\{XY^2,Y,Z\}$, $\{X,ZYZ^{-1},Z\}$, $\{X,Y,XZX^{-1}\}$, by $\varphi_1,\dots,\varphi_8$, respectively. Denote by $\mathfrak{a}$ the ideal in $\mathbb{Z}[x_1,\dots,x_{123}]$ generated by the functions $L_{ij}^{\varphi_m}$ where, as above, i,j stand for distinct numbers from the set $\{1,2,3\}$, and $k = 1,\dots,8$, and let*

$$Z_{\mathfrak{a}}(K) = \{x \in \mathbb{A}^7(K) : f(x) = 0 \text{ for all } f \in \mathfrak{a}\}.$$

Then for any field K of characteristic different from 2 we have

$$\begin{aligned} Z_{\mathfrak{a}}(K) = \{&(2,2,2,2,2,2,2),(0,-2,-2,0,0,2,0),(0,-2,2,0,0,-2,0),(0,2,-2,0,0,-2,0),\\ &(0,2,2,0,0,2,0),(0,0,0,-2,-2,-2,0),(0,0,0,-2,2,2,0),(0,0,0,2,-2,2,0),\\ &(0,0,0,2,2,-2,0)\}, \end{aligned}$$

and for any field of characteristic 2 we have $Z_{\mathfrak{a}}(K) = \{(0,0,0,0,0,0,0),(1,0,0,1,1,0,1)\}$.

Proof. MAGMA computation. □

For each of the points x appearing in Proposition 3.20 one can easily exhibit an explicit triple of matrices T such that $\pi(T) = x$. Say, $\pi(\mathrm{Id},\mathrm{Id},\mathrm{Id}) = (2,2,2,2,2,2,2)$,

$$\pi\left(\begin{pmatrix}0&-1\\1&0\end{pmatrix},\begin{pmatrix}-1&0\\0&-1\end{pmatrix},\begin{pmatrix}-1&0\\0&-1\end{pmatrix}\right) = (0,-2,-2,0,0,2,0),$$

and so on.

Theorem 3.10 is proved. □

Coming back to the map $\tilde{\varphi}$, let us consider an additional condition:

$$u = w(x,y), \tag{23}$$

where $x \in \widetilde{G}, y \in \widetilde{G}$ and $w \in F_2$. Let $g_3(a_1,a_2,a_{12}) = \mathrm{tr}(w(x,y))$, $g_{13}(a_1,a_2,a_{12}) = \mathrm{tr}(w(x,y)x)$, $g_{23}(a_1,a_2,a_{12}) = \mathrm{tr}(w(x,y)y)$, $g_{123}(a_1,a_2,a_{12}) = \mathrm{tr}(w(x,y)xy)$. Then (23) defines a three-dimensional variety $W(w) \subset Z$:

$$W(w) = Z \bigcap \left\{ \begin{array}{c} a_3 = g_3(a_1,a_2,a_{12}),\\ a_{13} = g_{13}(a_1,a_2,a_{12}),\\ a_{23} = g_{23}(a_1,a_2,a_{12}),\\ a_{123} = g_{123}(a_1,a_2,a_{12}) \end{array} \right\}. \tag{24}$$

We can now formulate a result which treats the $SL(2,q)$-case for three-variable maps and thus makes a crucial step towards getting a sufficient condition for a given sequence of type (5) to characterize finite solvable groups.

Theorem 3.21. *Let $v(x,u,y)$ and $w(x,y)$ be words in the free groups with three and two generators, respectively. Define a sequence $u_n(x,y)$ by the following recurrence relations:*

$$u_0(x,y) = w(x,y),\ \ u_{n+1}(x,y) = v(x,u_n(x,y),y).$$

Let $\varphi\colon \widetilde{G}\times\widetilde{G}\times\widetilde{G}\to\widetilde{G}$ be the map defined by $(x,u,y)\mapsto v(x,u,y)$, let $F(\varphi)$ be the variety of fixed points of the trace map ψ induced by φ (see diagram (14)), and let $W(w)$ be defined by (24). With the notation of Theorem 3.10, let $V=\{a_2=2, a_1=a_{12}, a_3=a_{23}, a_{13}=a_{123}\}$.

Assume that $F(\varphi)\bigcap W(w)$ contains a positive dimensional, absolutely irreducible $\mathbb{Q}$-subvariety Φ such that $\Phi' := \Phi\setminus(\Phi\bigcap V)$ is an open subset of Φ.

Then there is q_0 such that for every $q>q_0$ there exists a pair $(x,y)\in\widetilde{G}\times\widetilde{G}$ with $u_n(x,y)\neq 1$ for all $n\in\mathbb{N}$.

Proof. Let q_0 be such that $\Phi'(\mathbb{F}_q)\neq\emptyset$. Let $\overline{a}\in\Phi'(\mathbb{F}_q)$. By Theorem 3.10, there is a triple $(x,u,y)\in\widetilde{G}\times\widetilde{G}\times\widetilde{G}$ such that $\pi(x,u,y)=\overline{a}$. Moreover, since $\overline{a}\in W(w)$, we may take $u=w(x,y)$. Since $\overline{a}\in\Phi$, we have $\psi(\overline{a})=\overline{a}$, hence $\pi(x,u_1(x,y),y)=\overline{a}$. Similarly, $\pi(x,u_n(x,y),y)=\overline{a}$ for all $n\in\mathbb{N}$.

Since $a_2=\operatorname{tr}u_n(x,y)\neq 2$, we have $u_n(x,y)\neq 1$. □

Remark 3.22. Although this section was completely devoted to considering the group $SL(2)$ (until now $PSL(2)$ only appeared in its title), the obtained results (in particular, Theorems 3.6 and 3.21) are also applicable to the $PSL(2)$-case. (In the two-variable case, this is explicitly explained at the end of the proof of Theorem 3.6, the case of Theorem 3.21 is similar).

4. Case $G = Sz(q)$

In this section we consider a map $\varphi\colon G\times G\to G$ where G is a Suzuki group, $Sz(q)$, $q=2^{2m+1}$, $m\geq 1$. As above, for a fixed $y\in G$ we denote by $\varphi_y\colon G\to G$ the map $(x,y)\mapsto\varphi(x,y)$. There is no trace map in this case. Nevertheless there is a factorization (see diagram (27)) which simplifies the picture. This leads to a sufficient condition (Theorem 4.3) for the existence of periodic points. Although the condition is not that simple, we have an example in Subsection 5.1 when it works.

Recall that according to the Bruhat decomposition, $G=U_1\cup U_2$, where the first Bruhat cell $U_1=B$ consists of all lower-triangular matrices of the form $x=T(a,b)D(k)$ with

$$T(a,b)=\begin{pmatrix} 1 & 0 & 0 & 0\\ a & 1 & 0 & 0\\ a^{1+s}+b & a^s & 1 & 0\\ a^{2+s}+ab+b^s & b & a & 1\end{pmatrix},$$

$$D(k)=\begin{pmatrix} k^{s/2+1} & 0 & 0 & 0\\ 0 & k^{s/2} & 0 & 0\\ 0 & 0 & k^{-s/2} & 0\\ 0 & 0 & 0 & k^{-s/2-1}\end{pmatrix},$$

and the second Bruhat cell U_2 consists of the matrices

$$x=T(a,b)D(k)wT(c,d), \tag{25}$$

where

$$w=\begin{pmatrix} 0 & 0 & 0 & 1\\ 0 & 0 & 1 & 0\\ 0 & 1 & 0 & 0\\ 1 & 0 & 0 & 0\end{pmatrix}.$$

Here $a,b\in\mathbb{F}_q$, $k\in\mathbb{F}_q^*$, $s=2^{m+1}$.

Recall the following properties of these matrices:

(i) $T(0,1)T(a,b)=T(a,b)T(0,1)$;

(ii) $D(k)w = wD(k^{-1})$;
(iii) $T(a,b)T(c,d) = T(a+b, ac^s + b + d)$;
(iv) $wT(0,t)w = T(t^{1-s}, t^{-1})D(t^{2s/(s+2)})wT(t^{1-s}, 0)$;
(v) $T(0,1)^{-1} = T(0,1)$;
(vi) $D(k)^{-1}T(a,b)D(k) = T(ak, bk^{1+s})$.

For $x = T(a,b)D(k)wT(c,d) \in U_2$ define

$$x' = \varkappa(x) = T(c,d)xT(c,d)^{-1} = T(c,d)T(a,b)D(k)w = T(a+c, ca^s + b + d)D(k)w.$$

Note that for any $z = T(\alpha, \beta)$ we have

$$\begin{aligned}\varkappa(zxz^{-1}) &= \varkappa(T(\alpha,\beta)T(a,b)D(k)wT(c,d)T(\alpha,\beta)^{-1}) \\ &= T(c,d)T(\alpha,\beta)^{-1}T(\alpha,\beta)T(a,b)D(k)w = T(c,d)T(a,b)D(k)w = \varkappa(x).\end{aligned}$$

Lemma 4.1. *If for any $y, x, h \in G$ we have*

$$\varphi_y(hxh^{-1}) = h\varphi_{hyh^{-1}}(x)h^{-1}, \tag{26}$$

then for $y = T(0,t)$ we have

$$\varphi_y(\varkappa(x)) = \varkappa(\varphi_y(x)).$$

Proof. For $z = T(c,d)$ we have

$$\varphi_y(zxz^{-1}) = z\varphi_{z^{-1}yz}(x)z^{-1}.$$

Since the matrices $T(0,t)$ commute with any z, it follows that

$$\varphi_y(zxz^{-1}) = z\varphi_y(x)z^{-1},$$

i.e.

$$\varphi_y(\varkappa(x)) = \varkappa(z\varphi_y(x)z^{-1}) = \varkappa(\varphi_y(x)).$$

□

From now on until the end of this section we only consider elements x from the second Bruhat cell.

Corollary 4.2. *For $x \in U_2$ denote $\pi_1(x) = a+c$, $\pi_2(x) = ca^s+b+d$, $k(x) = k$. Then for $y = T(0,t)$ there exist functions f, g and h such that if $\varphi_y(x) \neq 1$ then*

$$\begin{aligned}\pi_1(\varphi_y(x)) &= f(\pi_1(x), \pi_2(x), k(x)), \\ \pi_2(\varphi_y(x)) &= g(\pi_1(x), \pi_2(x), k(x)), \\ k(\varphi_y(x)) &= h(\pi_1(x), \pi_2(x), k(x)).\end{aligned}$$

Proof. Indeed, by construction $\varkappa(x) = T(\pi_1(x), \pi_2(x))D(k(x))w$. Thus by Lemma 4.1, we have

$$T(\pi_1(\varphi_y(x)), \pi_2(\varphi_y(x))D(k(\varphi_y(x))w = \varkappa(\varphi_y(x)) = \varphi_y(\varkappa(x)) = \varphi_y(T(\pi_1(x), \pi_2(x))D(k(x))w).$$

It follows that $\pi_1(\varphi_y(x)), \pi_2(\varphi_y(x))$ and $k(\varphi_y(x))$ are determined uniquely by the values of $\pi_1(x)$, $\pi_2(x)$ and $k(x)$. □

Corollary 4.2 may be expressed by the following commutative diagram of $\mathbb{F}_q$-morphisms:

$$\begin{array}{ccc} \mathbb{A}^2_{a,b} \times \mathbb{A}^*_k \times \mathbb{A}^2_{c,d} \supseteq U & \xrightarrow{\varphi_y} & \mathbb{A}^2_{a,b} \times \mathbb{A}^*_k \times \mathbb{A}^2_{c,d} \\ \pi \big\downarrow & & \pi \big\downarrow \\ \mathbb{A}^2_{a,b} \times \mathbb{A}^*_k & \xrightarrow{\psi} & \mathbb{A}^2_{a,b} \times \mathbb{A}^*_k \end{array} \tag{27}$$

where U denotes the set of $x \in U_2$ such that $\varphi_y(x) \neq 1$.

This corollary provides the following sufficient condition for the existence of periodic points which can be viewed as an analogue of Theorem 3.6:

Theorem 4.3. *Let $G = Sz(q)$, let $y = T(0,1) \in G$, and suppose that the map φ_y satisfies the following conditions:*

- *equality (26) holds for any $x, y, h \in G$;*
- *the morphism $\psi\colon \mathbb{A}^2_{a,b} \times \mathbb{A}^*_k \to \mathbb{A}^2_{a,b} \times \mathbb{A}^*_k$ induced by φ_y (see diagram (27)) has an invariant set V (i.e. $\psi(V) \subset V$).*

Then the map $\varphi_y\colon G \to G$ has a non-identity periodic point.

Proof. Indeed, the cell U_2 does not contain the identity matrix. □

Remark 4.4. In view of (26), the statement of Theorem 4.3 holds for any $y = T(0,t)$.

5. Examples

In this section we want to demonstrate how the trace map works. In Subsection 5.1 we consider the two-variable case and give a new proof of the main theorem of [BWW] characterizing finite solvable groups. In Subsection 5.2 we compute the trace map for the three-variable sequence from [BGGKPP1], [BGGKPP2] (that also characterizes finite solvable groups). In Subsection 5.3 we apply our method for finding a modified sequence having the same property. Subsection 5.4 contains an illustration of the method for a simple case where the word under consideration is commutator.

5.1. **The sequence of Bray–Wilson–Wilson.** The sequence $s_n(x,y)$ of [BWW] is defined as follows:

$$s_1 = x, \quad s_2 = [y^{-1}xy, x^{-1}], \dots, s_n = [y^{-1}s_{n-1}y, s_{n-1}^{-1}], \dots,$$

Recall the main result of [BWW].

Theorem 5.1. *(*[BWW]*) A finite group G is solvable if and only if*

$$(\exists\ n \in \mathbb{N})\ (\forall (x,y) \in G \times G)\ s_n(x,y) = 1.$$

The proof reduces to the following:

Theorem 5.2. *(*[BWW]*) Let $G = PSL(2, \mathbb{F}_q)$, $q > 3$, or $G = Sz(2^{2m+1})$. Then there exists a pair $(x,y) \in G \times G$ such that $s_n(x,y) \neq 1$ for all $n \in \mathbb{N}$.*

We want to give another proof of Theorem 5.2 using the trace map and other geometric considerations.

For technical reasons we will change notation and consider a sequence $e_n(x,y)$ which differs from $s_n(x,y)$ only by replacing y with y^{-1}. Since in [BWW] the element y was supposed to be an involution, this does not matter. We define

$$e_1 = x, \quad e_2 = [yxy^{-1}, x^{-1}], \dots, e_n = [ye_{n-1}y^{-1}, e_{n-1}^{-1}], \dots,$$

i.e. in this example

$$\varphi(x,y) = \varphi_y(x) = [yxy^{-1}, x^{-1}]$$

(see Section 3).

Case of PSL As explained in Remark 3.22, we can freely apply the results of Subsection 3.1 obtained for $\widetilde{G} = SL(2,q)$ to the case $G = PSL(2,q)$.

We are going to compute the variety Φ of fixed points of the corresponding trace map $\psi\colon \mathbb{A}^3 \to \mathbb{A}^3$ (see diagram (8)). We maintain the notation of Subsection 3.1. In particular, we denote $s = \mathrm{tr}(x)$, $u = \mathrm{tr}(xy)$, $t = \mathrm{tr}(y)$, and $r = u^2 + s^2 + t^2 - ust$. Then (see [CMS, Lemma 5.2.4]),

$$f_1(s,u,t) = 2s^2 + (\mathrm{tr}(yxy^{-1}x^{-1}))^2 - s^2(\mathrm{tr}(yxy^{-1}x^{-1})) - 2,$$

$$\mathrm{tr}[y,x] = r - 2.$$

Direct computations give

$$f_1(s,u,t) = 2s^2 + (r-2)^2 - s^2(r-2) - 2 = s^2(4-r) + r^2 - 4r + 2 = (r-4)(r-s^2) + 2, \quad (28)$$

$$f_2(s,u,t) = f_1(s,u,t)\cdot t + s(st-u)(r-4) - t(r-3). \quad (29)$$

The variety $\Phi \subset \mathbb{A}^4$ is now defined by the following system:

$$\Phi = \left\{ \begin{array}{l} s = (r-4)(r-s^2) + 2, \\ u = st + s(st-u)(r-4) - t(r-3), \\ r = u^2 + t^2 + s^2 - ust. \end{array} \right\} \quad (30)$$

This curve contains a trivial component L_1:

$$s = 2, \qquad r = 4, \qquad u = t.$$

To eliminate this component, we consider a curve $\tilde{\Phi}$ in the space $\mathbb{A}^5$ with coordinates (s,u,t,r,z) which is isomorphic to $\Phi \setminus L_1$:

$$\tilde{\Phi} = \left\{ \begin{array}{l} r = u^2 + t^2 + s^2 - ust, \\ s = (r-4)(r-s^2) + 2, \\ u = st + s(st-u)(r-4) - t(r-3), \\ z(r-4) = 1. \end{array} \right\} \quad (31)$$

Lemma 5.3. *The plane curve $A \subset \mathbb{A}^2$ given by the equation $(s-2) = (r-4)(r-s^2)$ is a smooth irreducible genus 1 curve with two punctures.*

Proof. Assume the ground field is algebraically closed. Let $\tilde{A}$ be the closure of A in the projective space. One can check that $\tilde{A}$ has no singular points.

As a plane smooth curve, $\tilde{A}$ is irreducible. Moreover, it is a double cover of $\mathbb{P}^1$ and by Hurwitz's formula has genus 1. □

Magma computations show that the curve $\tilde{\Phi}$ has two components:

$$W_1 = \left\{ \begin{array}{l} z + t + s = 0, \\ u - t - s + r - 1 = 0, \\ ts - 2t - 2s + r = 0, \\ tr - 4t + sr - 4s + 1 = 0, \\ s^2 r - 4s^2 + s - r^2 + 4r - 2 = 0, \end{array} \right\} \quad (32)$$

$$W_2 = \left\{ \begin{array}{l} z - t + s = 0, \\ u - t + s - r + 1 = 0, \\ ts - 2t + 2s - r = 0, \\ tr - 4t - sr + 4s - 1 = 0, \\ s^2 r - 4s^2 + s - r^2 + 4r - 2 = 0, \end{array} \right\} \quad (33)$$

both defined over the ground field and isomorphic to $A \setminus \{r = 4, s = 2\}$, i.e. to a genus 1 irreducible curve with 3 punctures. Therefore both W_1 and W_2 are absolutely irreducible.

From Theorem 3.6 it follows that if q is big enough, then there exists a pair $(x,y) \in PSL(2,q) \times PSL(2,q)$ such that x is a periodic point of the map φ_y.

Remark 5.4. Since W_1, W_2 are curves of genus 1 with 3 punctures, they contain $\mathbb{F}_q$-points for all $q \geq 7$. Since each fibre contains a rational curve with at most two punctures, q "big enough" means $q \geq 7$ in this example. Small fields have been handled in a straightforward manner.

Case of $Sz(2^n)$

We keep the notation of Section 4. We have to show that the map φ_y meets all the conditions of Theorem 4.3. Condition (26) is obviously satisfied. Let us find an invariant set V of the map ψ (see diagram (27)). A direct computation of $f(0,b,k)$, $g(0,b,k)$ and $h(0,b,k)$ for $x = T(0,b)D(k)w$ and $y = T(0,1)$ gives

$$\begin{aligned} k(0,b,k) = k^2\beta^{\frac{2s}{s+2}} = k^2(b+1)^{\frac{2s}{s+2}} \cdot k^{\frac{(1+s)2s}{s+2}} = k^4(b+1)^{\frac{2s}{s+2}}, \\ f(0,b,k) = 0, \\ g(0,b,k) = (\beta^{1-s}k^{-1})^{1+s} + (\beta + 1/\beta)k^{-(1+s)} \\ = k^{-(1+s)}(\beta^{1-s^2} + \beta + 1/\beta) = k^{-(1+s)}\beta = b+1. \end{aligned}$$

Thus for $b \neq 0, 1$ the function g has period 2.

After the second iteration, we get

$$\begin{aligned} f(f(0,b,k), g(0,b,k), h(0,b,k)) = 0, \\ g(f(0,b,k), g(0,b,k), h(0,b,k)) = b, \\ h(f(0,b,k), g(0,b,k), h(0,b,k)) = k^{16}(b+1)^{\frac{8s}{s+2}}b^{\frac{2s}{s+2}}. \end{aligned}$$

Therefore, the set $V = \{x \in U_2 : \pi_1(x) = 0,\ \pi_2(x) = b \neq 0, 1\}$ is invariant under the second iteration of φ_y and does not contain 1.

Theorem 5.2 is proved.

5.2. **Three-variable sequence.** In this subsection we consider another sequence characterizing solvable groups which was introduced in [BGGKPP1], [BGGKPP2]:

$$u_0 = x^{-2}y^{-1}x, \ldots, u_{n+1} = [xu_nx^{-1}, yu_ny^{-1}], \ldots$$

In the notation of Subsection 3.2 we have

$$v(x,u,y) = [xux^{-1}, yuy^{-1}],\ w(x,y) = x^{-2}y^{-1}x,$$

and $\overline{a}$ stands for the point $\overline{a} = (a_1, a_2, a_3, a_{12}, a_{13}, a_{23}, a_{123}) \in \mathbb{A}^7$.

We need some additional notation:

$$\begin{aligned} a_{213} = \mathrm{tr}(yxu) = a_{12}a_3 + a_{13}a_2 + a_{23}a_1 - a_1a_2a_3, \\ b_{12} = \mathrm{tr}(x^{-1}y) = a_1a_2 - a_{12}, \\ b_{13} = \mathrm{tr}(x^{-1}u) = a_1a_3 - a_{13}, \\ b_{23} = \mathrm{tr}(y^{-1}u) = a_2a_3 - a_{23}, \\ b_{123} = \mathrm{tr}(x^{-1}yu) = a_1a_{23} - a_{123}, \\ b_{213} = \mathrm{tr}(y^{-1}xu) = a_2a_{13} - a_{213}, \\ c_{12} = \mathrm{tr}(xy^2) = a_{12}a_2 - a_1, \\ c_{m12} = \mathrm{tr}(x^{-1}y^2) = b_{12}a_2 - a_1, \\ d_{12} = \mathrm{tr}(x^2y) = a_{12}a_1 - a_2, \\ d_{m12} = \mathrm{tr}(x^{-2}y) = b_{12}a_1 - a_2, \\ g_{12} = \mathrm{tr}(xu^2) = a_{13}a_3 - a_1, \\ f_{m23} = \mathrm{tr}(u^2y^{-1}) = b_{23}a_3 - a_2, \\ p_1 = \mathrm{tr}(ux^{-1}yuy^{-1}x) = a_3b_{12}b_{123} - b_{12}^2 - b_{123}^2 + 2, \\ p_2 = b_{23}p_1 - b_{13}\{a_3b_{213} - b_{12}\} + a_1b_{213} - b_{23}, \\ p_3 = b_{12}(a_2p_1 - b_{13}b_{213} + d_{m12}) - b_{213}a_{23} + c_{m12}, \\ p_4 = b_{12}^2 + a_3^2 + b_{123}^2 - b_{12}a_3b_{123} - 2, \end{aligned}$$

$$p_5 = b_{12}^2 + a_3^2 + b_{213}^2 - b_{12}a_3b_{213} - 2,$$

$$l_1(\overline{a}) = 2a_3^2 + p_1^2 - p_1a_3^2 - 2,$$

$$l_2(\overline{a}) = a_1l_1 - b_{213}p_2 + p_3,$$

$$l_3(\overline{a}) = b_{213}(b_{13}p_1 - (b_{123}f_{m23} - b_{12}b_{23} + b_{13}))-$$
$$- b_{12}(p_1a_1 - b_{123}b_{23} + c_{m12}) + a_{13}b_{123} - d_{m12}.$$

A direct computation shows that

$$\mathrm{tr}([xux^{-1}, yuy^{-1}]) = l_1(\overline{a}), \tag{34}$$

$$\mathrm{tr}([xux^{-1}, yuy^{-1}]x) = l_2(\overline{a}), \tag{35}$$

$$\mathrm{tr}([xux^{-1}, yuy^{-1}]y) = l_3(\overline{a}). \tag{36}$$

In the following paragraph we compute

$$\mathrm{tr}([xux^{-1}, yuy^{-1}]xy) = l_4(\overline{a}) :$$

$Y = b_{13}b_{213} - d_{m12}$, $p_6 = b_{12}^2 + a_3^2 + b_{123}^2 - b_{12}a_3b_{123} - 2$, $G = b_{213}b_{12}a_3 - b_{12}^2 - b_{213}^2 + 2$, $U = a_2G - Y$, $V = b_{213}a_{23} - c_{m12}$, $E = b_{12}U - V$, $Q = b_{213}a_1 - b_{23}$, $R = a_3b_{213} - b_{12}$, $H = b_{13}R - Q$, $D = b_{23}G - H$, $B = b_{123}D - E$, $C = b_{12}(p_6 - 1)$, $A = a_2B - C$, $l_4 = a_{12}l_1 - A$.

Furthermore,

$$\mathrm{tr}(u_0) = \mathrm{tr}(x^{-2}y^{-1}x) = \mathrm{tr}(x^{-1}y^{-1}) = a_{12},$$

$$\mathrm{tr}(u_0x) = \mathrm{tr}(x^{-2}y^{-1}x^2) = \mathrm{tr}(y) = a_2,$$

$$\mathrm{tr}(u_0y) = \mathrm{tr}(x^{-2}y^{-1}xy) = \mathrm{tr}(x)\,\mathrm{tr}([x,y]) - \mathrm{tr}(y^{-1}xy) = a_1(a_1^2 + a_2^2 + a_{12}^2 - a_1a_2a_{12} - 3),$$

$$\mathrm{tr}(u_0xy) = \mathrm{tr}(x^{-2}y^{-1}x^2y) = \mathrm{tr}([x^2, y]) = (a_1 - 2)^2 + a_2^2 + d_{12}^2 - (a_1 - 2)a_2d_{12} - 2.$$

Therefore the variety $C = \Phi \bigcap W(w)$ is defined by equation (13) and the following system of equations:

$$l_1(\overline{a}) = a_3, \tag{37}$$

$$l_2(\overline{a}) = a_{13}, \tag{38}$$

$$l_3(\overline{a}) = a_{23}, \tag{39}$$

$$l_4(\overline{a}) = a_{123}, \tag{40}$$

$$a_3 = a_{12}, \tag{41}$$

$$a_{13} = a_2, \tag{42}$$

$$a_{23} = a_1(a_1^2 + a_2^2 + a_{12}^2 - a_1a_2a_{12} - 3), \tag{43}$$

$$a_{123} = (a_1 - 2)^2 + a_2^2 + d_{12}^2 - (a_1 - 2)a_2d_{12} - 2. \tag{44}$$

Magma computations show that C contains two components, C_1 and Φ: C_1 corresponds to the trivial solution $u_0 = 1$, $x = y^{-1}$, and Φ is an irreducible curve intersecting the set V (see Theorem 3.21 at a finite number of points (at most 31 as MAGMA computations give). Moreover, this curve is a projection of the solution of the equation $u_0 = u_1$ computed in [BGGKPP2].

5.3. **A new sequence.** In this subsection we produce a new sequence characterizing finite solvable groups. It is a modification of the sequence e_n considered in Subsection 5.1. We keep the notation of that subsection.

Let $\theta_n(x,y) = s_n(x,y^2)$. Denote $\theta(x,y) = \varphi(x,y^2)$, i.e.

$$\theta_y(x) = [y^2xy^{-2}, x^{-1}].$$

Theorem 5.5. *The map $\theta(x,y)\colon SL(2,q) \to SL(2,q)$ has nontrivial periodic points for all q.*

Proof. For a pair $(x,y) \in SL(2,q)$, let

$$s = \operatorname{tr}(x), \quad t_1 = \operatorname{tr}(y), \quad u_1 = \operatorname{tr}(xy), \quad t = \operatorname{tr}(y^2) = t_1^2 - 2, \qquad u = \operatorname{tr}(xy^2) = u_1t_1 - s.$$

Consider the following maps:

$$\varkappa : \mathbb{A}^3_{s,u_1,t_1} \longrightarrow \mathbb{A}^3_{s,u,t}, \qquad \varkappa(s,u_1,t_1) = (s, u_1t_1 - s, t_1^2 - 2);$$

$$\psi\colon \mathbb{A}^3_{s,u,t} \longrightarrow \mathbb{A}^3_{s,u,t}, \qquad \psi(s,u,t) = (f_1(s,u,t), f_2(s,u,t), t),$$

where the functions f_1 and f_2 are defined in (28) and (29), respectively;

$$\psi_\theta\colon \mathbb{A}^3_{s,u_1,t_1} \to \mathbb{A}^3_{s,u_1,t_1},$$

$$\psi_\theta(s,u_1,t_1) = (\operatorname{tr}\theta_y(x), \operatorname{tr}(\theta_y(x)\cdot y), \operatorname{tr} y).$$

We obtain the following commutative diagram:

$$\begin{array}{ccc} SL(2)\times SL(2) & \xrightarrow{(\theta,\mathrm{id})} & SL(2)\times SL(2) \\ \pi\downarrow & & \pi\downarrow \\ \mathbb{A}^3_{s,u_1,t_1} & \xrightarrow{\psi_\theta} & \mathbb{A}^3_{s,u_1,t_1} \\ \varkappa\downarrow & & \varkappa\downarrow \\ \mathbb{A}^3_{s,u,t} & \xrightarrow{\psi} & \mathbb{A}^3_{s,u,t} \end{array} \tag{45}$$

As shown above, the variety Φ of fixed points of ψ has three irreducible $\mathbb{F}_q$-components L_1, W_1, W_2, all absolutely irreducible for any q.

Lemma 5.6. *The curve $Z_2 := \varkappa^{-1}(W_2)$ is absolutely irreducible.*

Proof. Consider the curve $\overline{B}$ defined in $\mathbb{P}^3$ with homogeneous coordinates $(\tilde s : \tilde r : \tilde t : \tilde w)$ by the equations:

$$\tilde s\tilde t - 2\tilde t\tilde w + 2\tilde s\tilde w - \tilde r\tilde w = 0, \tag{46}$$

$$\tilde t\tilde r - 4\tilde t\tilde w - \tilde s\tilde r + 4\tilde s\tilde w - \tilde w^2 = 0, \tag{47}$$

$$(\tilde s - 2\tilde w)\tilde w^2 = (\tilde r\tilde w - \tilde s^2)(\tilde r - 4\tilde w). \tag{48}$$

Since equations (33) are linear in u and z, the curve $\overline{B}$ is isomorphic (or at least birational and one-to-one) to the projective closure of W_2.

The curve $\overline{C} \subset \mathbb{P}^4$, isomorphic (or at least birational and one-to-one) to the closure of Z_2, can be defined in $\mathbb{P}^4$ with coordinates $(\tilde t_1 : \tilde s : \tilde r : \tilde t : \tilde w)$ by the same system (46), (47), (48), together with the additional equation

$$\tilde t_1^2 = \tilde w(\tilde t + 2\tilde w). \tag{49}$$

The projection $\tau\colon \overline{C} \to \overline{B}$,

$$\tau(\tilde t_1 : \tilde s : \tilde r : \tilde t : \tilde w) = (\tilde s : \tilde r : \tilde t : \tilde w),$$

is a morphism which represents $\overline{C}$ as a ramified double cover of $\overline{B}$ (this can be checked by a direct computation). Since $\overline{B}$ is absolutely irreducible, so is $\overline{C}$. □

From diagram (45) it follows that at least the second iteration of ψ_θ has a nontrivial absolutely irreducible component in the variety of its fixed points. Formula (49) shows that $\overline{C}$ is a double cover of $\overline{B}$ with at most three ramification points (all at infinity). It follows that the genus is at most 2. Since B has 3 punctures and over at least one of them $\overline{C}$ is ramified, C has at most 5 punctures. Therefore for $q \geq 13$ there are points in Z_2 rational over $\mathbb{F}_q$.

The case $q < 13$ was checked by straightforward computations. □

Corollary 5.7. *A finite group G is solvable if and only if*

$$(\exists n \in \mathbb{N})\ (\forall (x,y) \in G \times G)\ \theta_n(x,y) = 1.$$

Proof. We argue as in the proof of Theorem 5.1. Theorem 5.5 settles the $PSL(2,q)$ case. In the case $Sz(2^n)$ no new proof is needed because $T(0,1) = T(1,1)^2$. Periodic points of φ_y with $y = T(0,1)$ are periodic points of θ_{y_1} with $y_1 = T(1,1)$. The case $G = PSL(3,3)$ is straightforward: for the matrices

$$x = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 2 \end{pmatrix}, \quad y = \begin{pmatrix} 0 & 2 & 2 \\ 1 & 2 & 1 \\ 0 & 2 & 0 \end{pmatrix}$$

we have $s_1(x,y) = s_4(x,y)$. □

Remark 5.8. The proof of [BWW] does not work for the sequence from Theorem 5.7. It is proved in [BWW] that for

$$y_0 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

there exists a periodic point of φ_{y_0} in $SL(2,q)$ for every q. But $y_0 \neq z^2$ in $SL(2,q)$ if 2 is not a square in $\mathbb{F}_q$.

Remark 5.9. We believe that the statement of Theorem 5.7 remains true if one takes y^n, with any $n \geq 2$, instead of y^2 (at least for even n) but this requires more subtle analysis.

5.4. **Commutator.** In the following example we want to show how useful the trace method can be. We present a very simple proof of the following statement (which is a very special case of a theorem of Borel [Bo], see also [La]):

Example 5.10. Let $G = SL(2,q)$. Then the map $F\colon G \times G \to G$ defined by $F(x,y) = [x,y]$ is a dominant morphism of the underlying algebraic $\mathbb{F}_q$-varieties.

Proof. In the notation of Subsection 3.1, consider the corresponding map $\psi\colon \mathbb{A}^3_{s,u,t} \to \mathbb{A}^3$: if $\mathrm{tr}(x) = s, \mathrm{tr}(y) = t, \mathrm{tr}(xy) = u$, then

$$\psi(s,u,t) = (f_1(s,u,t), f_2(s,u,t), t).$$

Here $f_1(s,u,t) = \mathrm{tr}(F(x,y)) = s^2 + t^2 + u^2 - ust - 2$, $f_2(s,u,t) = t$.

Let $z \in G$, and suppose that $a = \mathrm{tr}(z) \neq \pm 2$. We want to show that there exist $x, y \in G$ with $[x,y] = z$.

For any $t \in \mathbb{F}_q$ consider the inverse image $\Gamma_{a,t,t} := \psi^{-1}(a,t,t) \subset \mathbb{A}^3_{s,u,t}$. We have

$$\Gamma_{a,t,t} = \{(s,u,t) \in \mathbb{A}^3 :\ s^2 + t^2 + u^2 - ust - 2 - a = 0\}.$$

For a fixed value $t_0 \neq \pm 2$, this is a quadratic equation in (s,u) which has a solution (s_0,u_0) over every finite field (cf. the proof of Theorem 3.4). Thus we have a point $Q := (s_0,u_0,t_0) \in \Gamma_{a,t_0,t_0}$.

By Theorem 3.4, $\pi^{-1}(Q) \neq \emptyset$, so take $(x,y) \in \pi^{-1}(Q)$. We have $\mathrm{tr}(F(x,y)) = a = \mathrm{tr}(z)$. If $a \neq \pm 2$ (i.e. z is semisimple), $F(x,y)$ is conjugate to z, i.e. $[x,y] = wzw^{-1}$. We get $[w^{-1}xw, w^{-1}yw] = z$, as required. □

The map $F\colon G\times G\to G$ provides a dynamical system on $SL(2,q)\times SL(2,q)$ with $\tilde{\phi}(x,y)=([x,y],y)$. It corresponds to the Engel sequence $e_1=[x,y],\dots,\ e_{n+1}=[e_n,y],\dots$

Let us show that this dynamical system has nontrivial periodic points for every q. The cases $q=2,3$ are treated by a direct computation, so assume $q>3$. In view of Theorem 3.4, it is sufficient to find a fixed point of the trace map

$$\psi(s,u,t)=(s^2+t^2+u^2-ust-2,t,t)$$

with $s^2\neq 4$, $t^2\neq 4$. The point (s,t,t) is fixed if $s=s^2+2t^2-st^2-2$.If $q=2^n$, then any pair $(s=1+t^2,t)$ is a needed solution of this equation. If $q\neq 2^n$, then for a fixed t we get $s_1=2$ (forbidden), $s_2=t^2-1$. Thus, any pair (t^2-1,t), $t^2\neq -1,3,4$ provides a needed fixed point.

6. Possible generalizations

Here we present some more general problems arising from concrete calculations of the preceding sections. In Subsection 6.1 we consider AG systems introduced in Section 1 making this notion more precise. In particular, we want to distinguish between the cases when the underlying geometric object is defined over a global field or its ring of integers. We define residually periodic dynamical systems, propose some relevant conjectures and give several examples. In Subsection 6.2 we discuss in more detail verbal dynamical systems defined in the introduction. By combining the notions of AG dynamical system and verbal dynamical system, we define systems carrying both structures.

6.1. **Residually periodic dynamical systems.**

We start with AG dynamical systems.

Let K be a global field, and let $\mathcal{O}$ stand for the ring of integers in K.

Definition 6.1. A triple $D=(X,V,\varphi)$ is called a K-dynamical system if

- X is an algebraic K-variety;
- $\varphi\colon X\to X$ is a dominant K-morphism;
- $V\subset X(K)$ is a φ-invariant subset.

Definition 6.2. A triple $\mathcal{D}=(\mathcal{X},\mathcal{V},\Phi)$ is called an $\mathcal{O}$-dynamical system if

- $\mathcal{X}$ is an $\mathcal{O}$-scheme of finite type;
- $\Phi\colon\mathcal{X}\to\mathcal{X}$ is a dominant $\mathcal{O}$-morphism;
- $\mathcal{V}\subset\mathcal{X}(\mathcal{O})$ is a Φ-invariant subset.

We say that an $\mathcal{O}$-dynamical system $\mathcal{D}=(\mathcal{X},\mathcal{V},\Phi)$ is an integral model of $D=(X,V,\varphi)$ if

- $\mathcal{X}\times_{\mathcal{O}}K=X$;
- the restriction of Φ to the generic fibre coincides with φ;
- $R(\mathcal{V})=V$, where $R\colon\mathcal{X}\to X$ is the restriction to the generic fibre.

Consider a K-dynamical system $D=(X,V,\varphi)$ and its integral model $\mathcal{D}=(\mathcal{X},\mathcal{V},\Phi)$. For a place p of K let

- $\varkappa_p$ be the residue field of p;
- $\mathcal{X}_p$ the special fibre of $\mathcal{X}$ at p;
- $R_p\colon\mathcal{X}\to\mathcal{X}_p$ the reduction map (restriction to the special fibre);
- $\varphi_p\colon\mathcal{X}_p\to\mathcal{X}_p$ the reduction of Φ viewed as a morphism of $\varkappa_p$-schemes;
- $X_p=\mathcal{X}_p(\varkappa_p)$ the set of rational points;
- $V_p=R_p(\mathcal{V})\subset X_p$ the reduction of $\mathcal{V}$.

Assume that for all but finitely many places p the scheme $\mathcal{X}_p$ is integral. One can deduce from [Gr, 9.6.1(ii)] that for all but finitely many p's the reduced morphism φ_p is dominant. Let $z\in X_p\setminus V_p$ be a periodic point of φ_p. Let $\ell(z)$ be the number of distinct points in the orbit of z. Set $\ell_p:=\min\{\ell(z)\}$ where the minimum is taken over all z's as above. If there are no periodic points in $X_p\setminus V_p$, we set $\ell_p=\infty$. Let M denote the collection of primes p such that $\ell_p=\infty$. Let $N=\{\ell_p\}_{p\notin M}$.

Definition 6.3. With the above notation, we say that a K-dynamical system $D = (X, V, \varphi)$ or an $\mathcal{O}$-dynamical system $\mathcal{D} = (\mathcal{X}, \mathcal{V}, \Phi)$ is *residually aperiodic* if the set M is infinite, *residually periodic* if M is finite, and *strongly residually periodic* if the sets M and N are both finite.

For example, in Subsection 5.1 for a map $\psi\colon \mathbb{A}^3_{\mathbb{Z}} \to \mathbb{A}^3_{\mathbb{Z}}$ we had $\mathcal{X} = \mathbb{A}^3$, $\varkappa_p = \mathbb{F}_p$, $\mathcal{V} = \{(\pm 2, \pm t, t)\}$, $N = \{1\}$ and $M = \emptyset$.

We believe that the following special case is particularly interesting. Let $\mathcal{V} \subset \mathcal{X}(\mathcal{O})$ be the set of all preperiodic integer points (i.e. points having a finite orbit). Let $V_p = R_p(\mathcal{V}) \subset X_p$ be its reduction mod p. Residual periodicity of $\mathcal{D} = (\mathcal{X}, \mathcal{V}, \Phi)$ means that φ_p has periodic points outside V_p for all but finitely many p's. In simple words, we are looking for periodic points of φ_p not coming from preperiodic integer points of Φ. Note that according to [Si2], cycles coming from a fixed nonperiodic integer point cannot be too short (their length, as a function of the cardinality of the residue field, tends to infinity). Thus our approach to studying cycles of reduced systems is, in a sense, complementary to [Si2].

As mentioned in the introduction, there may be different reasons for a dynamical system to be residually periodic. For higher-dimensional systems one can look for geometric conditions. The next notion captures the phenomenon of extra coordinates, or more generally invariant functions, as in Example 1.8.

Definition 6.4. We say that a dynamical system $D = (X, V, \varphi)$ is of *fibred* type if there exists a regular function f on X such that $f \circ \varphi^{(n)} = f$ for some iteration $\varphi^{(n)}$ of φ.

Question 6.5. Assume that a dynamical system $D = (X, V, \varphi)$ is of *fibred* type. Assume that the endomorphism φ is not birational. Under what conditions on φ is D strongly residually periodic?

Question 6.5 is essentially higher-dimensional. In one-dimensional situations the main role, of course, belongs to arithmetic. To get a better feeling of the problem, it is useful to consider one-dimensional examples which are, in a sense, opposite to Example 1.7 from the introduction.

Example 6.6. Let $\mathcal{T} = \mathbb{G}_{\mathrm{m},\mathbb{Z}} = \mathrm{Spec}(\mathbb{Z}[x, y]/(xy - 1))$ be the trivial one-dimensional torus. Fix a positive integer d, and let $\Phi\colon \mathcal{T} \to \mathcal{T}$ denote the power map: $t \to t^d$. The set of integer points $\mathcal{R} = \mathcal{T}(\mathbb{Z})$ consists of two points, 1 and -1, both fixed under Φ (i.e. periodic with period one). We choose the forbidden set $\mathcal{V} = \mathcal{R}$. If $d = 2$, then $\ell_p = \infty$ for every Fermat prime $p = 2^m + 1$. Thus the system is residually periodic or aperiodic depending on whether there are finitely or infinitely many Fermat primes. Assume now that d is odd.

Proposition 6.7. *The dynamical system $(\mathcal{T}, \mathcal{R}, \Phi)$ of Example 6.6 is residually periodic but is not strongly residually periodic.*

Proof. We have $X_p = \mathbb{F}_p^*$, and for any $t \in \mathbb{F}_p^*$ we have $\varphi^{(n)}(t) = t^{d^n}$. Assume $(p, d) = 1$. We are looking for $t \neq \pm 1$ such that

$$t^{d^n - 1} \equiv 1 \pmod{p}. \tag{50}$$

To find such a t with minimal possible n, let us first introduce some notation. For any prime ℓ such that $(d, \ell) = 1$ denote by s_ℓ the order of d in $\mathbb{F}_\ell^*$. Denote by $Q(p) = \{q_i\}$ the set of all odd primes appearing in the prime decomposition of $p - 1$ and coprime to d. Set $a(p) := \min_{q \in Q} s_q$. If $p \equiv 1 \pmod 4$, we have $\ell_p \leq 2$. We claim that for $p \equiv -1 \pmod 4$ we have $\ell_p = a(p)$. Indeed, suppose that the minimum is achieved at some $q \in Q$, so $d^{s_q} - 1 = qm$ for some integer m. If g is a primitive element of $\mathbb{F}_p$, one can take $t = g^{(p-1)/q}$ and $n = s_q$ to satisfy (50). On the other hand, if $n < s_q$, then by the definition of s_q we have $n < s_\ell$ for all $\ell \in Q$, and hence for all such ℓ we have

$$d^n \not\equiv 1 \pmod{\ell}.$$

The above also holds for all ℓ dividing d, so we conclude that $(d^n - 1, p - 1) = 2$. If (50) holds for some t, then the order of t must divide both $d^n - 1$ and $p - 1$, hence it is equal to 2. Thus $t = -1$

and belongs to the reduction of the forbidden set $\mathcal{R}$. We conclude that (50) does not hold for any $n < s_q$. This means that $s_q = a(p)$ is the minimal possible length of the orbit of φ_p, i.e. $\ell_p = a(p)$.

To finish the proof of the proposition, it is enough to establish the following simple lemma (we thank Z. Rudnick for an elementary proof):

Lemma 6.8. *The set $A = \{a(p)\}$, where p runs over all prime numbers congruent to -1 modulo 4, is infinite.*

Proof of the Lemma. Assume the contrary:

$$A = \{s_{q_1}, \dots, s_{q_t}\}. \tag{51}$$

To get a contradiction, we wish to find $p \equiv -1 \pmod 4$ with $a(p) \notin A$.

First note that there are at most finitely primes q with a given value of s_q, and denote by B the set of all q such that $s_q \in A$. It follows that B is finite. Thus we have to find a prime p such that $p - 1$ is not divisible by any $q \in B$. We want to find a prime number p satisfying the system of congruences

$$x \equiv -1 \pmod 4,$$
$$x \equiv -1 \pmod q$$

for all $q \in B$. By the Chinese Remainder Theorem, the solutions of this system form an arithmetic progression. By Dirichlet's Prime Number Theorem, this progression contains infinitely many primes. If now p is such a prime, we have $p \not\equiv 1 \pmod q$ for any $q \in B$. Thus the order of d in $\mathbb{F}_p^*$ is not equal to any of s_{q_i}'s, and so $p \notin A$, contradiction.

This finishes the proof of the lemma and hence of Proposition 6.7. □

Example 6.9. Let now E be a CM elliptic curve defined over $\mathbb{Q}$ by the equation $y^2 = x^3 - x$, and let $\mathcal{E}$ denote its minimal Weierstrass model. Let $\Phi\colon \mathcal{E} \to \mathcal{E}$ be the multiplication-by-d map (d stands for a positive odd integer). There are four 2-torsion points: (0,0), (1,0), (-1,0) and ∞, all belonging to $\mathcal{E}(\mathbb{Z})$. Denote this collection by $\mathcal{V}$. If $p \equiv -1 \pmod 4$, the reduction of E is supersingular, i.e. $|E(\mathbb{F}_p)| = p + 1$. We can now denote by $b(p)$ the smallest prime factor of the number $|E(\mathbb{F}_p)|/4$ and by the argument similar to that of the previous example show that the set $B = \{b(p)\}$, where p runs over all $p \equiv -1 \pmod 4$, is infinite. This leads to

Proposition 6.10. *The dynamical system $\mathcal{D} = (\mathcal{E}, \mathcal{V}, \Phi)$ is residually periodic but is not strongly residually periodic.* □

The interested reader is invited to complete the details of the proof as well as to develop more examples of arithmetical interest.

To go beyond CM elliptic curves, one needs more efforts. A natural question to ask is the following one:

Question 6.11. Let E be an elliptic curve over $\mathbb{Q}$, and let D denote the order of its rational torsion. For each place p of good reduction, denote by $c(p)$ the smallest prime divisor of the number $|E(\mathbb{F}_p)|/D$. Can the set $C = \{c(p)\}$, where p runs over all places of good reduction of E, be finite? Can the system $(\mathcal{E}, \mathcal{E}(\mathbb{Q})_{tors}, \Phi)$ be strongly residually periodic?

At the first glance, the conjectures by Lang–Trotter [LT] and Koblitz [Ko], predicting (for most elliptic curves) infinitely many p's with $|E(\mathbb{F}_p)|$ of prime order, give little hope to find an example of an elliptic curve such that the dynamical system defined by the multiplication-by-d map is strongly residually periodic. However, the following example (due to N. Jones) prevents from making hasty conclusions. Consider the curve E_0 given over $\mathbb{Q}$ by the Weierstrass equation

$$y^2 = x^3 + 75x + 125.$$

N. Jones proved that although E_0 has no rational torsion, the order of $E_0(\mathbb{F}_p)$ is divisible either by 2 or by 3 for all $p > 5$. The curve E_0 is of Mordell–Weil rank 1, so the multiplication-by-d map

Φ induces a nontrivial dynamical system $\mathcal{D} = (\mathcal{E}_0, \infty, \Phi)$. Taking, say, $d = 7$, we conclude that $\mathcal{D}$ is strongly residually periodic in the strongest possible sense: it has no periodic points but the residual system $\mathcal{D}_p$ has a fixed point for all $p > 5$ (compare with Example 1.7).

On the other hand, N. Jones proved (unconditionally on Koblitz's conjectures) that for a "typical" elliptic curve E over $\mathbb{Q}$ an analogue of Lemma 6.8 indeed holds which implies that the dynamical system $\mathcal{D}$ is not strongly residually periodic for such an E, i.e. typically the answer to Question 6.11 is negative. See the Appendix for more details.

6.2. **Verbal dynamical systems on group schemes.** We view the calculations of Section 3 as a first step in attacking one of the most important conceptual questions left open after discovery of two-variable sequences characterizing finite solvable groups: for a sequence of words in the free group on two generators, to what extent the property to characterize the class of finite solvable groups is a property of general position, and what type of the dynamic behaviour is typical? Questions of such "nonbinary" type, which do not admit an answer of type "yes-no", have been considered by many mathematicians, from Poincaré to Arnold, as the most interesting ones. Dynamics of word maps in free group, in spirit of [LP], [La], [Sh], [LS], [GS], led to a breakthrough in some classical problems of the theory of finite groups, and it may happen to play a crucial role in the above mentioned problem as well. Namely, a possible goal is to prove (or disprove) that for a sufficiently wide class of sequences the property to characterize the class of finite solvable groups holds in "general position" and is determined by its dynamics in the free group. In what follows F_r stands for the free group on r generators.

Question 6.12. Suppose that a sequence $\overrightarrow{u} = u_1, u_2, \dots, u_n, \dots$ of elements of F_2 satisfies the following conditions:

(i) $u_n(a, 1) = u_n(1, g) = 1$ for all sufficiently big n, every group G, and all elements $a, g \in G$;

(ii) if G is any group and a, g are elements of G such that $u_n(a, g) = 1$, then for every $m > n$ we have $u_m(a, g) = 1$;

(iii) no element of $\overrightarrow{u}$ equals 1 in F_2;

(iv) there exists N such that for all $n > N$ the word $u_n(x, y)$ belongs to the n-th derived subgroup $F_2^{(n)}$ of F_2.

Is it true that if a finite group G satisfies an identity $u_n(x, y) \equiv 1$ for some n, then it is solvable?

In connection with Question 6.12, it is natural to pose

Problem 6.13. To describe the words in F_2 satisfying conditions (i)–(iv) of Question 6.12.

Extensive MAGMA computations show strong numerical evidence of a positive answer to Question 6.12, at least for the class of sequences $\overrightarrow{u}$ studied in [Ri]: $u_0 := f, \dots, u_n := [gu_ng^{-1}, hu_nh^{-1}], \dots$, where f, g, h stand for some words from F_2.

One can put Question 6.12 into somewhat more general context. Towards this end, we suggest to combine the notions of verbal and AG dynamical systems defined in Section 1. For simplicity we restrict ourselves to considering $\mathbb{Z}$-dynamical systems.

Definition 6.14. A verbal dynamical $\mathbb{Z}$-system consists of the following setup:

- positive integers r, s;
- an r-tuple $\mathcal{W} = (w_1, \dots, w_r)$ of words in the free group F_{r+s};
- an r-tuple $\mathcal{J} = (f_1, \dots, f_s)$ of words in the free group F_s (optional);
- a group scheme $\mathcal{G}$ of finite type over $\mathbb{Z}$;
- a set $I \subset \mathcal{G}^{r+s}(\mathbb{Z})$.

The following assumptions are to be satisfied.

(i) Let $D_{\mathcal{W}} \colon \mathcal{G}^{r+s} \to \mathcal{G}^{r+s}$ be a morphism of $\mathbb{Z}$-schemes defined on the group $G = \mathcal{G}^{r+s}(A)$ of A-points of $\mathcal{G}^{r+s}$ for every $\mathbb{Z}$-algebra A by the formula

$$(g_1, \dots, g_s, v_1, \dots, v_r) \mapsto (g_1, \dots, g_s, w_1(g_1, \dots, g_s, v_1, \dots, v_r), \dots, w_r(g_1, \dots, g_s, v_1, \dots, v_r).$$

Then we assume that $D_{\mathcal{W}}$ is *dominant*.

(ii) The set I is *invariant*, i.e. $D_{\mathcal{W}}(I) \subset I$.

Our earlier considerations (cf. Examples 1.2 and 1.4) naturally fit into this setting if $\mathcal{G}$ is a semisimple Chevalley group scheme over $\mathbb{Z}$ (e.g., $\mathcal{G} = SL(2,\mathbb{Z})$ as in the present paper). Indeed, in that case by a theorem of Borel ([Bo], see also [La]), the morphism $D_{\mathcal{W}}$ is dominant, and we arrive at a verbal dynamical $\mathbb{Z}$-system in the sense of Definition 6.14. Remark 3.8 shows that the dynamical systems on $SL(2,p)$ relevant for our original problem, can be viewed as special fibres of the original verbal $\mathbb{Z}$-system.

Remark 6.15. It would be interesting to formulate a word-theoretic condition on $\mathcal{W}$ guaranteeing that for any Chevalley group scheme $\mathcal{G}$ the morphism $D_{\mathcal{W}}$ is dominant.

In connection with Question 6.5 one can pose

Problem 6.16. Given a verbal dynamical $\mathbb{Z}$-system, that is not of fibred type, find conditions under which it is (strongly) residually periodic.

In particular, it would be interesting to consider the system from Section 3.1 given by the map $\varphi_y\colon SL(2,\mathbb{Z}) \to SL(2,\mathbb{Z})$ (y fixed) with $I = \{1\}$. This system has an invariant rational function, but it is not regular. It was proven in [BWW] that for

$$y = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

it is residually periodic. On the other hand, our numerical experiments give some evidence that it is not strongly residually periodic.

We believe that verbal dynamical systems deserve more thorough study. To the best of our knowledge, most arithmetically interesting questions, in spirit of the monograph [Si1] (boundedness of periods, distributions of periods in reductions, various local-global problems), are widely open (or even not yet posed at all).

Acknowledgements. Bandman and Kunyavskiĭ were supported in part by the Ministry of Absorption (Israel) and the Israel Academy of Sciences grant 1178/06. A substantial part of this work was done during the visits of Bandman and Kunyavskiĭ to MPIM (Bonn) in 2007 and 2009 and the visits of Grunewald to Bar-Ilan University in 2008 and the Hebrew University of Jerusalem in 2009, and discussed by all the coauthors during the international workshops (2007, 2009) hosted by the Heinrich-Heine-Universität (Düsseldorf), and the Oberwolfach meeting "Profinite and Geometric Group Theory" in 2008 (the visits were supported in part by the Minerva Foundation through the Emmy Noether Research Institute of Mathematics). The appendix to the paper arose from questions posed by Kunyavskiĭ to Jones when they participated in the program "Diophantine equations" organized by the Hausdorff Research Institute for Mathematics (Bonn) in 2009. The support of all above institutions is highly appreciated.

We are very grateful to V. Berkovich, F. Campana, J.-L. Colliot-Thélène, N. Fakhruddin, M. Leyenson, Z. Rudnick, M. Tyomkin, and S. Vishkautsan for useful discussions and correspondence.

References

[BGGKPP1] T. Bandman, G.-M. Greuel, F. Grunewald, B. Kunyavskiĭ, G. Pfister, and E. Plotkin, Two-variable identities for finite solvable groups, *C.R. Acad. Sci. Paris, Ser. I* **337** (2003), 581–586.

[BGGKPP2] T. Bandman, G.-M. Greuel, F. Grunewald, B. Kunyavskiĭ, G. Pfister, and E. Plotkin, Identities for finite solvable groups and equations in finite simple groups, *Compositio Math.* **142** (2006), 734–764.

[BB] D. Berend and Y. Bilu, Polynomials with roots modulo every integer, *Proc. Amer. Math. Soc.* **124** (1996), 1663–1671.

[Bo] A. Borel, On free subgroups of semisimple groups, *Enseign. Math.* **29** (1983), 151–164.

[Br] R. Brandl, Integer polynomials with roots mod p for all primes p, *J. Algebra* **204** (2001), 822–835.

[BBH] R. Brandl, D. Bubboloni, and I. Hupp, Polynomials with roots mod p for all primes p, *J. Group Theory* **4** (2001), 233–239.

[BWW] J. N. Bray, J. S. Wilson, and R. A. Wilson, Characterization of finite soluble groups by laws in two variables, *Bull. London Math. Soc.* **37** (2005), 179–186.

[CMS] J. Cossey, Sh. Oates Macdonald, and A. P. Street, On the laws of certain finite groups, *J. Australian Math. Soc.* **11** (1970), 441–489.

[Fr] R. Fricke, Über die Theorie der automorphen Modulgruppen, *Nachr. Akad. Wiss. Göttingen* (1896), 91–101.

[FK] R. Fricke and F. Klein, *Vorlesungen der automorphen Funktionen*, vol. 1–2, Teubner, Leipzig 1897, 1912.

[FJ] M. D. Fried and M. Jarden, *Field arithmetic*, 3rd ed., Springer-Verlag, Berlin 2008.

[GS] S. Garion and A. Shalev, Commutator maps, measure preservation, and T-systems, *Trans. Amer. Math. Soc.* **361** (2009), 4631–4651.

[Go] W. Goldman, An exposition of results of Fricke and Vogt. Preprint, available at http://www.math.umd.edu/˜wmg/publications.html .

[Gr] A. Grothendieck (with collaboration of J. Dieudonné), Eléments de Géométrie Algébrique IV, Étude locale des schémas et des morphismes de schémas (troisième partie), *Inst. Hautes Etudes Sci. Publ. Math.* **28** (1966).

[Ho] R. D. Horowitz, Characters of free groups represented in the two-dimensional special linear group, *Comm. Pure Appl. Math.* **25** (1972), 635–649.

[Hr] E. Hrushovski, The elementary theory of the Frobenius automorphisms. Preprint arXiv:math/LO/0406514.

[HB] B. Huppert and N. Blackburn, *Finite groups*, III, Springer-Verlag, Berlin–Heidelberg–New York 1982.

[Ko] N. Koblitz, Primality of the number of points on an elliptic curve over a finite field, *Pacific J. Math.* **131** (1988), 157–165.

[LT] S. Lang and H. Trotter, *Frobenius distributions in GL_2-extensions*, Lecture Notes Math. **504**, Springer-Verlag, Berlin–New York 1976.

[LW] S. Lang and A. Weil, Number of points of varieties in finite fields, *Amer. J. Math.* **76** (1954), 819–827.

[La] M. Larsen, Word maps have large image, *Israel J. Math.* **139** (2004), 149–156.

[LP] M. Larsen and R. Pink, Finite subgroups of algebraic groups. Preprint, 1999.

[LS] M. Larsen and A. Shalev, Word maps and Waring type problems, *J. Amer. Math. Soc.* **22** (2009), 437–466.

[Ma1] W. Magnus, Rings of Fricke characters and automorphisms groups of free groups, *Math. Z.* **170** (1980), 91–102. In *Collected papers*, Springer-Verlag, New York, 1984, 687–699.

[Ma2] W. Magnus, The uses of 2 by 2 matrices in combinatorial group theory. A survey, *Resultate Math.* **4** (1981), 171–192. In *Collected papers*, Springer-Verlag, New York, 1984, 701–722.

[Na] R. Narkiewicz, Polynomial cycles in algebraic number fields, *Colloq. Math.* **58** (1989), 151–155.

[Pe1] J. Peyrière, On an article by W. Magnus on the Fricke characters of free groups, *J. Algebra* **228** (2000), 659–673.

[Pe2] J. Peyrière, Polynomial dynamical systems associated with substitutions. In *Substitutions in dynamics, arithmetic and combinatorics*, Lecture Notes Math. **1794**, Springer, Berlin 2002, 321–343.

[Ri] E. Ribnere, Sequences of words characterizing finite solvable groups, *Monatshefte Math.* **157** (2009), 387–401.

[Sh] A. Shalev, Word maps, conjugacy classes, and a non-commutative Waring-type theorem, *Ann. Math.*, to appear.

[Si1] J. H. Silverman, *The arithmetic of dynamical systems*, Springer-Verlag, New York 2007.

[Si2] J. H. Silverman, Variation of periods modulo p in arithmetic dynamics, *New York J. Math.* **14** (2008), 601–616.

[So] J. Sonn, Polynomials with roots in $\mathbb{Q}_p$ for all p, *Proc. Amer. Math. Soc.* **136** (2008), 1955–1960.

[Th] J. Thompson, Non-solvable finite groups all of whose local subgroups are solvable, *Bull. Amer. Math. Soc.* **74** (1968), 383–437.

[Vo] H. Vogt, Sur les invariants fundamentaux des equations différentielles linéaires du second ordre, *Ann. Sci. E.N.S, 3-ième Sér.* **4** (1889), Suppl. S.3–S.70.

Appendix. Primes p for which $\#E(\mathbb{F}_p)$ has only large prime factors

Nathan Jones

A1. Introduction

Let E be an elliptic curve over $\mathbb{Q}$ of conductor N_E. For each prime p of good reduction for E, consider the group $E(\mathbb{F}_p)$ of $\mathbb{F}_p$-points of E. In 1988, Koblitz [3] conjectured a precise asymptotic formula for the number of good primes p up to x for which $\#E(\mathbb{F}_p)$ is prime.

Conjecture A1. *There exists a precise constant $\mathfrak{S}_E \geq 0$ so that*

$$\#\{p \leq x : p \nmid N_E \text{ and } \#E(\mathbb{F}_p) \text{ is prime}\} = \mathfrak{S}_E \cdot \frac{x}{\log^2 x} + o\left(\frac{x}{\log^2 x}\right),$$

as $x \longrightarrow \infty$.

In particular, provided the constant $\mathfrak{S}_E > 0$, Conjecture A1 implies that there are infinitely many primes p for which $\#E(\mathbb{F}_p)$ is prime. In case $\mathfrak{S}_E = 0$, one can prove (as a consequence of the Chebotarev density theorem) that $\#E(\mathbb{F}_p)$ is prime for only finitely many primes p.

Based on the precise form of the predicted constant $\mathfrak{S}_E$, Koblitz further noted that $\mathfrak{S}_E$ is positive if and only if every other elliptic curve E' over $\mathbb{Q}$ which is $\mathbb{Q}$-isogenous to E has no non-trivial rational torsion:

$$\mathfrak{S}_E > 0 \Longleftrightarrow \left(E' \sim_{\mathbb{Q}} E \Rightarrow E'(\mathbb{Q})_{\text{tors}} = \{\mathcal{O}_{E'}\}\right). \tag{A-1}$$

However, because of a technical error in the underlying heuristic, the constant $\mathfrak{S}_E$ appearing in the original conjecture is incorrect. A refined conjecture, which in particular corrects $\mathfrak{S}_E$, has recently been given by D. Zywina [8]. In the interest of consistency, let us henceforth understand the symbol $\mathfrak{S}_E$ appearing in Conjecture A1 to refer to the corrected constant $C_{E,1}$ appearing in [8, Conjecture 1.2] (we will describe this constant in more detail in Section A2). Having thus replaced $\mathfrak{S}_E$, the interpretation (A-1) of exactly when $\mathfrak{S}_E$ is positive is no longer valid. We will show this in Section A4 by exhibiting an elliptic curve E over $\mathbb{Q}$ for which the right-hand side of (A-1) is true, but for which $\mathfrak{S}_E = 0$ nevertheless.

In spite of various partial results (see for instance [1] and the references therein), Conjecture A1 is still open. Our goal is to prove the following theorem, wherein we relax "is prime" to "has only large prime factors." Let us denote by

$$c_E(p) := \min\{\ell \text{ prime} : \ell \mid \#E(\mathbb{F}_p)\}$$

the smallest prime divisor of $\#E(\mathbb{F}_p)$.

Theorem A2. *Suppose that*

$$\mathfrak{S}_E > 0,$$

where $\mathfrak{S}_E$ is the constant appearing in Conjecture A1. Then the set

$$\{c_E(p) : p \nmid N_E\}$$

is unbounded.

In other words, Theorem A2 asserts that, for each $x > 0$, there exists a prime number $p = p(E, x)$ such that for any prime number ℓ we have

$$\ell \mid \#E(\mathbb{F}_p) \Longrightarrow \ell > x.$$

We remark that one could likely prove something stronger by employing the appropriate tools. In the interest of brevity and simplicity, we content ourselves with Theorem A2.

We will begin by describing precisely the constant $\mathfrak{S}_E$, from which it will be evident that the converse of Theorem A2 holds, i.e. for any elliptic curve E over $\mathbb{Q}$, one has

$$\mathfrak{S}_E = 0 \implies \{c_E(p) : p \nmid N_E\} \text{ is bounded.} \tag{A-2}$$

We will then prove Theorem A2. Finally, we will discuss the issue of exactly when one has $\mathfrak{S}_E > 0$ and give an example of an elliptic curve E over $\mathbb{Q}$ for which $\mathfrak{S}_E = 0$ (and for which $\{c_E(p) : p \nmid N_E\}$ is bounded, thus illustrating (A-2)). Throughout, ℓ and p will always denote prime numbers.

A2. The heuristic of Conjecture A1 and the constant $\mathfrak{S}_E$

The heuristic leading to Conjecture A1 is analogous to the one which leads to the classical twin prime conjecture (see [3] and [8] for more details), and changes slightly depending on whether or not E has complex multiplication (CM). As usual, for $p \nmid N_E$, define the integer $a_E(p)$ by the formula

$$\#E(\mathbb{F}_p) =: p + 1 - a_E(p). \tag{A-3}$$

By a theorem due originally to Hasse, we have that $|a_E(p)| \leq 2\sqrt{p}$, and so the size of $\#E(\mathbb{F}_p)$ is near the size of p. Thus, regarding p and $\#E(\mathbb{F}_p)$ as two independently chosen random positive integers of size x, the "probability" that they are both prime should satisfy

$$\mathcal{P}(p \text{ is prime and } \#E(\mathbb{F}_p) \text{ is prime}) \approx \frac{1}{(\log x)^2}, \tag{A-4}$$

by the prime number theorem. However, this prediction fails to take into account arithmetic information about the reductions of p and $\#E(\mathbb{F}_p)$ modulo positive integers. In order to describe how one corrects the situation, we begin by recalling the division fields attached to E and Chebotarev density theorem.

A2.1. **The division fields $\mathbb{Q}(E[n])$ of E.** For each positive integer $n \geq 1$ denote by

$$E[n] := \{P \in E(\overline{\mathbb{Q}}) : [n](P) = \mathcal{O}_E\}$$

the n-torsion of E and by $\mathbb{Q}(E[n])$ the n-th division field of E, i.e. the field generated by the x and y coordinates of each $P \in E[n]$. The field $\mathbb{Q}(E[n])$ is a Galois extension of $\mathbb{Q}$, and by fixing a $\mathbb{Z}/n\mathbb{Z}$-basis of $E[n]$, we may (and henceforth will) view $\mathrm{Gal}\,(\mathbb{Q}(E[n])/\mathbb{Q})$ as a subgroup of $GL_2(\mathbb{Z}/n\mathbb{Z})$:

$$\mathrm{Gal}\,(\mathbb{Q}(E[n])/\mathbb{Q}) \subseteq GL_2(\mathbb{Z}/n\mathbb{Z}).$$

The following proposition, which relates p and $a_E(p)$ with $\mathbb{Q}(E[n])$ is well-known. In its statement $\sigma_{\mathbb{Q}(E[n])/\mathbb{Q}}(p) \subset \mathrm{Gal}\,(\mathbb{Q}(E[n])/\mathbb{Q}) \subseteq GL_2(\mathbb{Z}/n\mathbb{Z})$ denotes the conjugacy class of a Frobenius automorphism at p, which we view as a subset of $GL_2(\mathbb{Z}/n\mathbb{Z})$.

Proposition A3. *For any positive integer n and any prime p of good reduction for E which does not divide n, p is unramified in $\mathbb{Q}(E[n])$. Furthermore,*

$$\mathrm{tr}(\sigma_{\mathbb{Q}(E[n])/\mathbb{Q}}(p)) \equiv a_E(p) \mod n$$

and

$$\det(\sigma_{\mathbb{Q}(E[n])/\mathbb{Q}}(p)) \equiv p \mod n.$$

A2.2. **The Chebotarev density theorem.** Recall the Chebotarev density theorem [7]. Let L/F be a (finite) Galois extension of number fields and $\mathcal{C} \subseteq \mathrm{Gal}\,(L/F)$ any subset which is stable by $\mathrm{Gal}\,(L/F)$-conjugation. Denote by Σ_F the set of prime ideals of F and

$$\Sigma_F(x) := \{\mathfrak{p} \in \Sigma_F : N_{F/\mathbb{Q}}(\mathfrak{p}) \leq x\}.$$

For each prime ideal $\mathfrak{p} \in \Sigma_F$ which is unramified in L, let $\sigma_{L/F}(\mathfrak{p}) \subseteq \mathrm{Gal}\,(L/F)$ denote the conjugacy class of the Frobenius element attached to any prime $\mathfrak{P}$ of L lying over $\mathfrak{p}$.

Theorem A4. *(The Chebotarev density theorem) We have*

$$\lim_{x\to\infty} \frac{\#\{\mathfrak{p} \in \Sigma_F(x) : \mathfrak{p} \text{ unramified in } L \text{ and } \sigma_{L/F}(\mathfrak{p}) \subseteq \mathcal{C}\}}{\#\Sigma_F(x)} = \frac{\#\mathcal{C}}{\#Gal\,(L/F)}.$$

In probabilistic terms, Theorem A4 says that the probability that a randomly selected prime ideal $\mathfrak{p}$ satisfies $\sigma_{L/K}(\mathfrak{p}) \subseteq \mathcal{C}$ is $\frac{\#\mathcal{C}}{\#\mathrm{Gal}\,(L/F)}$.

A2.3. **Correcting the naive heuristic** (A-4)**.** For any positive integer n and subgroup $G \leq GL_2(\mathbb{Z}/n\mathbb{Z})$, define the subset $\Omega_n(G) \subseteq G$ by

$$\Omega_n(G) := \{g \in G : \det(g) + 1 - \mathrm{tr}(g) \notin (\mathbb{Z}/n\mathbb{Z})^*\}. \tag{A-5}$$

The probability that a large randomly chosen integer is coprime to n is $\dfrac{\#(\mathbb{Z}/n\mathbb{Z})^*}{\#(\mathbb{Z}/n\mathbb{Z})}$. On the other hand, by (A-3), Proposition A3 and Theorem A4, the probability that $\#E(\mathbb{F}_p)$ is coprime with n is

$$\frac{\#(\mathrm{Gal}\,(\mathbb{Q}(E[n])/\mathbb{Q}) - \Omega_n(\mathrm{Gal}\,(\mathbb{Q}(E[n])/\mathbb{Q})))}{\#(\mathrm{Gal}\,(\mathbb{Q}(E[n])/\mathbb{Q}))}.$$

Thus, it is natural to multiply (A-4) by the correction factor

$$\frac{\dfrac{\#(\mathrm{Gal}\,(\mathbb{Q}(E[n])/\mathbb{Q}) - \Omega_n(\mathrm{Gal}\,(\mathbb{Q}(E[n])/\mathbb{Q})))}{\#(\mathrm{Gal}\,(\mathbb{Q}(E[n])/\mathbb{Q}))}}{\dfrac{\#(\mathbb{Z}/n\mathbb{Z})^*}{\#(\mathbb{Z}/n\mathbb{Z})}}. \tag{A-6}$$

Noting that

$$\Omega_n(\mathrm{Gal}\,(\mathbb{Q}(E[n])/\mathbb{Q})) = \pi^{-1}\left(\Omega_{\delta(n)}\left(\mathrm{Gal}\,(\mathbb{Q}(E[\delta(n)])/\mathbb{Q})\right)\right),$$

where $\delta(n) := \prod_{\ell|n} \ell$ denotes the square-free kernel of n and $\pi : GL_2(\mathbb{Z}/n\mathbb{Z}) \twoheadrightarrow GL_2(\mathbb{Z}/\delta(n)\mathbb{Z})$ denotes the canonical projection, we see that (A-6) only depends on $\delta(n)$, and so it suffices to consider square-free n. Defining

$$n = n(z) := \prod_{\ell \leq z} \ell \tag{A-7}$$

to be the square-free number supported on primes $\ell \leq z$, we multiply (A-4) by (A-6) and take the limit as $z \to \infty$, arriving at Conjecture 1 with

$$\mathfrak{S}_E := \lim_{z\to\infty} \frac{\left(1 - \dfrac{\#\Omega_{n(z)}(\mathrm{Gal}\,(\mathbb{Q}(E[n(z)])/\mathbb{Q}))}{\#\mathrm{Gal}\,(\mathbb{Q}(E[n(z)])/\mathbb{Q})}\right)}{\displaystyle\prod_{\ell|n(z)} (1 - 1/\ell)}. \tag{A-8}$$

Our next proposition describes $\mathfrak{S}_E$ in more detail. In particular, it implies that the limit in (A-8) converges to a finite positive limit, provided it is non-zero for each fixed $z \geq 2$.

Proposition A5. *Let E be an elliptic curve over $\mathbb{Q}$ and let $\mathfrak{S}_E$ be defined by* (A-8)*. There exists a positive square-free integer $n_E \geq 1$ and a real number $\lambda_E > 0$ so that*

$$\mathfrak{S}_E = \frac{\left(1 - \dfrac{\#\Omega_{n_E}(Gal\,(\mathbb{Q}(E[n_E])/\mathbb{Q}))}{\#Gal\,(\mathbb{Q}(E[n_E])/\mathbb{Q})}\right)}{\displaystyle\prod_{\ell|n_E} (1 - 1/\ell)} \cdot \lambda_E.$$

Proof. In the CM case, this follows from [4, Corollaire, p. 302] and in the non-CM case from [4, (2), p. 260]. For more details, see [8]. □

Although it won't be necessary in what follows, we remark that

$$\lambda_E = \begin{cases} \dfrac{1}{2} \cdot \displaystyle\prod_{\ell \nmid n_E} \left(1 - \chi(\ell)\frac{\ell^2 - \ell - 1}{(\ell - \chi(\ell))(\ell - 1)^2}\right) & \text{if } E \text{ has CM by } K, \\ \displaystyle\prod_{\ell \nmid n_E} \left(1 - \frac{\ell^2 - \ell - 1}{(\ell - 1)^3(\ell + 1)}\right) & \text{if } E \text{ has no CM,} \end{cases}$$

where in the CM case, $\chi(\ell) \in \{\pm 1\}$ denotes the Kronecker symbol giving the splitting of ℓ in the imaginary quadratic field K.

Corollary A6. *We have*

$$\mathfrak{S}_E = 0 \iff (\exists \text{ square-free } n_0,\ \Omega_{n_0}(Gal(\mathbb{Q}(E[n_0])/\mathbb{Q})) = Gal(\mathbb{Q}(E[n_0])/\mathbb{Q}))\,.$$

In particular, if $\mathfrak{S}_E = 0$, then by (A-3), Proposition A3 and Theorem A4, we have

$$p \nmid n_0 \cdot N_E \implies \gcd(\#E(\mathbb{F}_p), n_0) > 1. \tag{A-9}$$

Since this in turn causes $\{c_E(p) : p \nmid N_E\}$ to be bounded, we have verified (A-2).

A3. Proof of Theorem A2

To prove Theorem A2, we will apply Theorem A4 with $F = \mathbb{Q}$, $L = \mathbb{Q}(E[n])$, and

$$\mathcal{C} = (\mathrm{Gal}\,(\mathbb{Q}(E[n])/\mathbb{Q}) - \Omega_n(\mathrm{Gal}\,(\mathbb{Q}(E[n])/\mathbb{Q}))\,,$$

with $\Omega_n(G)$ as in (A-5) and $n = n(z)$ as in (A-7). Fix any prime $p > z$ which doesn't divide N_E. By Proposition A3, p is unramified in $\mathbb{Q}(E[n(z)])$ and furthermore we have the following equivalence:

$$(\forall \ell \le z,\ \ell \nmid \#E(\mathbb{F}_p)) \iff \sigma_{\mathbb{Q}(E[n(z)])/\mathbb{Q}}(p) \nsubseteq \Omega_{n(z)}(\mathrm{Gal}\,(\mathbb{Q}(E[n(z)])/\mathbb{Q})). \tag{A-10}$$

Now consider the Chebotarev factor

$$\mathcal{D}_z := \frac{\#(\mathrm{Gal}\,(\mathbb{Q}(E[n(z)])/\mathbb{Q}) - \Omega_{n(z)}(\mathrm{Gal}\,(\mathbb{Q}(E[n(z)])/\mathbb{Q})))}{\#(\mathrm{Gal}\,(\mathbb{Q}(E[n(z)])/\mathbb{Q}))}.$$

By Corollary A6, we see that

$$\mathfrak{S}_E > 0 \implies \mathcal{D}_z > 0.$$

Thus, provided $\mathfrak{S}_E > 0$, Theorem A4 implies the existence of a prime number $p_1 = p_1(E, z)$ for which

$$\sigma_{\mathbb{Q}(E[n(z)])/\mathbb{Q}}(p_1) \nsubseteq \Omega_{n(z)}(\mathrm{Gal}\,(\mathbb{Q}(E[n(z)])/\mathbb{Q})).$$

By (A-10), we see that for each $\ell \le z$, ℓ does not divide $\#E(\mathbb{F}_{p_1})$, and so $c_E(p_1) \ge z$. Since z was arbitrary, Theorem A2 follows.

A4. The positivity of $\mathfrak{S}_E$

It is now natural to ask: under what conditions is the constant $\mathfrak{S}_E$ positive? Because of the Weil Pairing (see [6], for example), for any level n, we have that the determinant map restricts to a surjection

$$\det\colon \mathrm{Gal}\,(\mathbb{Q}(E[n(z)])/\mathbb{Q}) \twoheadrightarrow (\mathbb{Z}/n(z)\mathbb{Z})^*.$$

By Corollary A6, we are thus led to ask the following question.

Question A7. Let $n \ge 1$ be a positive square-free integer, and let $G \le GL_2(\mathbb{Z}/n\mathbb{Z})$ be a subgroup for which the determinant map restricts to a surjection:

$$\det\colon G \twoheadrightarrow (\mathbb{Z}/n\mathbb{Z})^*.$$

Under which circumstances do we have $\Omega_n(G) = G$?

It is clear from the definitions that, for any ℓ dividing n we have

$$\Omega_\ell(G \mod \ell) = G \mod \ell \quad \implies \quad \Omega_n(G) = G.$$

We join Serre [4, I-2] in leaving the following exercise up to the reader.

Exercise A8. Prove that, for any subgroup $G_\ell \le GL_2(\mathbb{Z}/\ell\mathbb{Z})$, $\Omega(G_\ell) = G_\ell$ if and only if either

$$G_\ell \subseteq \left\{ \begin{pmatrix} 1 & * \\ 0 & * \end{pmatrix} \right\} \qquad \text{or} \qquad G_\ell \subseteq \left\{ \begin{pmatrix} * & * \\ 0 & 1 \end{pmatrix} \right\}.$$

Furthermore, $\mathrm{Gal}\,(\mathbb{Q}(E[\ell])/\mathbb{Q}) = G_\ell$ as above if and only if E is isogenous over $\mathbb{Q}$ to some elliptic curve E' over $\mathbb{Q}$ satisfying $E'[\ell](\mathbb{Q}) \neq \{\mathcal{O}_{E'}\}$ (in the first case, E' is simply E). We record this as

Remark A9. If E is $\mathbb{Q}$-isogenous to some elliptic curve E' over $\mathbb{Q}$ for which $E'(\mathbb{Q})_{\mathrm{tors}} \neq \{\mathcal{O}_{E'}\}$, then $\mathfrak{S}_E = 0$.

It is tempting to expect (as Koblitz did) that the converse of Remark A9 also holds, but the following example shows that this is not the case. Let $\ell \neq 2$ be any prime and consider the subgroup $G \leq GL_2(\mathbb{Z}/2\mathbb{Z}) \times GL_2(\mathbb{Z}/\ell\mathbb{Z})$ defined by

$$G = \{(g_2, g_\ell) \in GL_2(\mathbb{Z}/2\mathbb{Z}) \times G_1(\mathbb{Z}/\ell\mathbb{Z}) : \chi_2(g_2) = \chi_\ell(g_\ell)\}, \tag{A-11}$$

where

$$G_1(\mathbb{Z}/\ell\mathbb{Z}) := \left\{ \begin{pmatrix} \pm 1 & * \\ 0 & * \end{pmatrix} \right\} \leq GL_2(\mathbb{Z}/\ell\mathbb{Z}) \tag{A-12}$$

and the characters χ_2 and χ_ℓ are defined by

$$\chi_2 : GL_2(\mathbb{Z}/2\mathbb{Z}) \longrightarrow GL_2(\mathbb{Z}/2\mathbb{Z})/[GL_2(\mathbb{Z}/2\mathbb{Z}), GL_2(\mathbb{Z}/2\mathbb{Z})] \simeq \{\pm 1\}$$

and

$$\chi_\ell \left(\begin{pmatrix} \pm 1 & * \\ 0 & * \end{pmatrix} \right) = \pm 1. \tag{A-13}$$

Notice that, even though

$$\Omega_2(G \bmod 2) \subsetneq G \bmod 2 \quad \text{and} \quad \Omega_\ell(G \bmod \ell) \subsetneq G \bmod \ell,$$

we have $\Omega_{2\ell}(G) = G$. Provided we can find an elliptic curve E over $\mathbb{Q}$ with $\mathrm{Gal}\,(\mathbb{Q}(E[2\ell])/\mathbb{Q}) \leq G$, then $\#E(\mathbb{F}_p)$ will only be prime finitely often because whenever it is not divisible by 2, it must be divisible by ℓ, and vice versa.

A4.1. **A counterexample to** (A-1)**.**

Proposition A10. *Let E be the elliptic curve defined by the Weierstrass equation*

$$y^2 = x^3 + 75x + 125.$$

For any elliptic curve E' over $\mathbb{Q}$ which is $\mathbb{Q}$-isogenous to E, one has $E'(\mathbb{Q})_{tors} = \{\mathcal{O}_{E'}\}$. Nevertheless, $\mathfrak{S}_E = 0$. Furthermore, the Mordell–Weil group attached to E is infinite:

$$\#E(\mathbb{Q}) = \infty.$$

Proof. Since $N_E = 2^2 \cdot 3^3 \cdot 5^2$, we see that E has good reduction away from $p \in \{2, 3, 5\}$. One calculates that $\#E(\mathbb{F}_7) = 4$ and $\#E(\mathbb{F}_{17}) = 21$, from which it follows that, for any E' over $\mathbb{Q}$ which is $\mathbb{Q}$-isogenous to E, we have $E'(\mathbb{Q})_{\mathrm{tors}} = \{\mathcal{O}_{E'}\}$. On the other hand, recall that $\mathbb{Q}(E[2]) = \mathbb{Q}(\text{the roots of } x^3 + 75x + 125)$, so that $\sqrt{\Delta_E} = 2^2 \cdot 3 \cdot 5^3\sqrt{-15} \in \mathbb{Q}(E[2])$. Also, the point $(-5, 5\sqrt{-15}) \in E[3](\mathbb{Q}(\sqrt{-15}))$ shows that

$$\mathbb{Q}(\sqrt{\Delta_E}) = \mathbb{Q}(\sqrt{-15}) \subseteq \mathbb{Q}(E[2]) \bigcap \mathbb{Q}(E[3]).$$

It follows that, taking $\ell = 3$ in (A-11), we have $\mathrm{Gal}\,(\mathbb{Q}(E[6])/\mathbb{Q}) \leq G$, where χ_2 and χ_ℓ correspond to the restriction map

$$\mathrm{Gal}\,(\mathbb{Q}(E[6])/\mathbb{Q}) \longrightarrow \mathrm{Gal}\,(\mathbb{Q}(\sqrt{\Delta_E})/\mathbb{Q}) \simeq \{\pm 1\}.$$

Taking $n_0 = 6$ in Corollary A6, we see that $\mathfrak{S}_E = 0$.

Finally, the point $(5, 25) \in E(\mathbb{Q})$ is of infinite order, and so $\#E(\mathbb{Q}) = \infty$, as claimed. □

Furthermore, one can readily verify (A-9) with $n_0 = 6$ and E as in Proposition A10, as follows. For any rational prime $p \geq 7$ and choice of Frobenius automorphism $\sigma_6(p) \in \sigma_{\mathbb{Q}(E[6])/\mathbb{Q}}(p)$, we have that

$$\sigma_6(p)(\sqrt{\Delta_E}) = \sqrt{\Delta_E} \Rightarrow \sigma_{\mathbb{Q}(E[3])/\mathbb{Q}}(p) \subseteq \Omega_3(\mathrm{Gal}\,(\mathbb{Q}(E[3])/\mathbb{Q})) \Rightarrow 3 \mid \#E(\mathbb{F}_p)$$

and

$$\sigma_6(p)(\sqrt{\Delta_E}) = -\sqrt{\Delta_E} \Rightarrow \sigma_{\mathbb{Q}(E[2])/\mathbb{Q}}(p) \subseteq \Omega_2(\mathrm{Gal}\,(\mathbb{Q}(E[2])/\mathbb{Q})) \Rightarrow 2 \mid \#E(\mathbb{F}_p).$$

Since $\sqrt{\Delta_E} = 2^2 \cdot 3 \cdot 5^3\sqrt{-15}$, it follows that for $p \nmid 30$, we have

$$\left(\frac{-15}{p}\right) = 1 \Rightarrow 3 \mid \#E(\mathbb{F}_p)$$

and

$$\left(\frac{-15}{p}\right) = -1 \Rightarrow 2 \mid \#E(\mathbb{F}_p).$$

This verifies (A-9) and shows that

$$\{c_E(p) : p \nmid N_E\} = \{2, 3\}.$$

More generally, we have

Remark A11. If E is $\mathbb{Q}$-isogenous to some elliptic curve E' over $\mathbb{Q}$ for which $E'(\mathbb{Q}(\sqrt{\Delta_{E'}}))_{\mathrm{tors}} \neq \{\mathcal{O}_{E'}\}$, then $\mathfrak{S}_E = 0$.

Have we covered all possible cases where $\mathfrak{S}_E = 0$? We will now give an example of a subgroup $G \leq GL_2(\mathbb{Z}/3\ell\mathbb{Z})$ satisfying $\Omega_{3\ell}(G) = G$, where $\ell \geq 5$ is some prime. Let

$$\mathcal{N}_3 := \left\{\pm\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \pm\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}\right\} \sqcup \left\{\pm\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \pm\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\right\} \leq GL_2(\mathbb{Z}/3\mathbb{Z}),$$

and define

$$G := \{(g_3, g_\ell) \in \mathcal{N}_3 \times G_1(\mathbb{Z}/\ell\mathbb{Z}) : \det g_3 = \chi_\ell(g_\ell)\},$$

where $G_1(\mathbb{Z}/\ell\mathbb{Z})$ and χ_ℓ are as in (A-12) and (A-13), respectively, and we are regarding $\det(g_3) \in \mathbb{F}_3^* = \{\pm 1\}$. As before, we have

$$\Omega_3(G \mod 3) \subsetneq G \mod 3 \quad \text{and} \quad \Omega_\ell(G \mod \ell) \subsetneq G \mod \ell,$$

but $\Omega_{3\ell}(G) = G$. Perhaps there may also be an elliptic curve E over $\mathbb{Q}$ with $\mathrm{Gal}\,(\mathbb{Q}(E[3\ell])/\mathbb{Q}) \leq G$, though we haven't explicitly exhibited one.

A4.2. **Serre curves.** A **Serre curve** is an elliptic curve E over $\mathbb{Q}$ for which

$$\forall n \geq 1, \quad [GL_2(\mathbb{Z}/n\mathbb{Z}) : G_E(n)] \leq 2.$$

(Intuitively, a Serre curve is an elliptic curve for which $\mathrm{Gal}\,(\mathbb{Q}(E[n])/\mathbb{Q})$ is "as large as possible" for each $n \geq 1$.) We remark that, as shown in [8, Proposition 4.2], we have

$$E \text{ is a Serre curve } \implies \mathfrak{S}_E > 0.$$

When ordered according to naive height, almost all elliptic curves are Serre curves (see [2]). Thus, for a "typical" elliptic curve E over $\mathbb{Q}$ one has $\mathfrak{S}_E > 0$.

A5. Concluding remarks

As mentioned in the introduction, one can likely prove stronger forms of Theorem A2. For instance, one could probably use an effective version of the Chebotarev density theorem to obtain a quantitative upper bound for the smallest prime p for which $c_E(p) > x$.

Since we have not completely resolved it, we record here

Question A12. Under what conditions do we have $\mathfrak{S}_E > 0$?

The examples discussed in Section A4 seem to indicate that this question is more delicate than it first may seem. Conjecture A1 has also been generalized to the context where E is defined over a general number field K (see [8]), in which case the answer to Question A12 may become even more delicate.

References to the appendix

[1] A. Balog, A. C. Cojocaru, and C. David, Average twin prime conjecture for elliptic curves. Preprint, available at http://www.mathstat.concordia.ca/faculty/cdavid/publi.html .
[2] N. Jones, Almost all elliptic curves are Serre curves, *Trans. Amer. Math. Soc.*, to appear.
[3] N. Koblitz, Primality of the number of points on an elliptic curve over a finite field, *Pacific J. Math.* **131** (1988), 157–165.
[4] J-P. Serre, *Abelian l-adic representations and elliptic curves*, Benjamin, New York–Amsterdam 1968.
[5] J-P. Serre, Propriétés galoisiennes des points d'ordre fini des courbes elliptiques, *Invent. Math.* **15** (1972), 259–331.
[6] J. H. Silverman, *The arithmetic of elliptic curves*, Springer-Verlag, New York 1986.
[7] N. Tschebotareff, Die Bestimmungder Dichtigkeit einer Menge von Primzahlen, welche zu einer gegebenen Substitutionsklasse gehören, *Math. Ann.* **95** (1926), 191–228.
[8] D. Zywina, A refinement of Koblitz's conjecture. Preprint, available at http://www.math.upenn.edu/~zywina/ .

Bandman: Department of Mathematics, Bar-Ilan University, 52900 Ramat Gan, ISRAEL
E-mail address: bandman@macs.biu.ac.il

Grunewald: Mathematisches Institut der Heinrich-Heine-Universität Düsseldorf, Universitätsstr. 1, 40225 Düsseldorf, GERMANY
E-mail address: grunewald@math.uni-duesseldorf.de

Kunyavskii: Department of Mathematics, Bar-Ilan University, 52900 Ramat Gan, ISRAEL
E-mail address: kunyav@macs.biu.ac.il

Jones: Department of Mathematics, University of Mississippi, Hume Hall 305, P. O. Box 1848, University, MS 38677-1848, USA
E-mail address: ncjones@olemiss.edu